

Where now for NASA?

The loss of the space shuttle Columbia and its seven astronauts has cast a shadow over America's space agency. But it also presents an opportunity to redirect the US space programme towards the lofty goal of exploration.

Three times in the space age, Americans have placed flowers on the graves of astronauts, but under such different circumstances that they might have happened in three different countries. The Apollo 1 launchpad fire in 1967 was a tragic setback, but it failed to deter a determined nation in its race to reach the Moon. The 1986 Challenger disaster, perhaps the most unsettling of the three, struck hard at the psyche of a less confident people. And now we have Columbia: coming between the terrorist outrages of 11 September 2001 and a looming war in the Middle East, it feels sadly familiar, the latest in a series of grim developments.

In the downtime forced by the accident investigation, it is appropriate to ask where NASA should go from here. After the Apollo and Challenger accidents, both of which came early in their respective programmes, it was clear that the answer — once the safety issues had been addressed — was to keep working towards much the same goals as before. But the space shuttles have been flying for more than two decades now, and the loss of Columbia comes at a time when NASA was already rethinking its approach to human space flight.

Last autumn, the agency unveiled a new space transportation plan that calls for a small Orbital Space Plane to be built. Launched on a conventional rocket, it would provide an alternative means of reaching space. The plan comes after a series of failed attempts to develop a larger, fully re-usable space plane, and after NASA bailed out on a mini-shuttle concept that it had been developing as a 'lifeboat' for the International Space Station.

Critics charge that the Orbital Space Plane will take so long to develop that it could serve little useful purpose with regard to the space station, and that NASA should have gone ahead with its mini-shuttle instead. But at least it represents an attempt to get on with building a new space vehicle for the first time in decades, instead of simply producing artists' concepts.

When NASA administrator Sean O'Keefe took over the space agency in December 2001, he was criticized for lacking vision and focusing on accounting. Yet the budget for 2004 that he unveiled on Monday (see page 565) — without much fanfare, given the sombre mood of the moment — includes the seeds of a vision that could see astronauts venturing beyond the confines of Earth orbit.

Bold vision

With the blessing of the White House, NASA has revived the old idea of nuclear-powered rockets, which most people in the space business agree is the only way to push ahead with ambitious space exploration. For years, discussion of nuclear propulsion was forbidden at NASA, for fear of a negative public reaction. But O'Keefe's agency seems unafraid of this debate.

Also downplayed in recent years was the possibility of sending astronauts beyond Earth orbit. There have always been a few engineers, tucked away in the deepest corners of NASA centres, who were allowed to think about missions to the Moon, Mars and elsewhere. But their work was outside the mainstream, and wasn't attached to any larger programme. In this week's budget proposal, NASA has requested \$39 million for a new Human Research

Initiative, with a planned budget of \$347 million over five years. Its unabashed goal is to find safe ways for people to spend extended periods in space.

By themselves, such moves are hardly enough to accomplish the objective of sending people beyond Earth orbit. But marking them out as specific initiatives, for everyone to see, is an important first step towards this goal.

Nature has been critical of human space flight in the past, particularly when it relies spuriously on science to justify projects that have difficulty standing on their own merits. If the Human Research Initiative isn't linked to longer-term plans for human space exploration, then this same objection will hold, as it would be difficult to see any point to the endeavour other than continuing to justify the space station's existence.

NASA might therefore seize the moment, and honour its fallen astronauts, by further encouraging its visionaries to come out from their hiding places. They have plenty of ideas that could be executed in the next decade or so. One is to send astronauts as field geologists to investigate near-Earth asteroids. Another is to have them service telescopes at the L2 lagrangian point, more than one-and-a-half million kilometres from Earth, where the successors to the Hubble Space Telescope will be located (see *Nature* **419**, 666; 2002). In the longer term, of course, Mars beckons.

Common goal

Sending astronauts beyond Earth orbit again will be tremendously expensive. But if America's political leaders do decide that exploration should be NASA's goal, then all of the agency's workforce could work towards a common purpose, much as it did in the Apollo era. As it stands now, space scientists and engineers developing concepts for future crewed spacecraft have little to do with each other.

Space scientists who use orbiting observatories to study the cosmos or send robotic probes to explore the planets might be wary of any such coming together. It is they, after all, who have been doing most of the real exploring since the Apollo programme. And after seeing many of their projects delayed after the Challenger accident, they might be reluctant to hitch their fortunes to human space flight once again.

There are positive examples of cooperation between NASA's two 'camps', however. Notable among these are the Hubble repair missions, which first salvaged the telescope after it was launched with a faulty mirror, and which have subsequently continued to upgrade this historic scientific asset.

The immediate aftermath of the Columbia tragedy, which reminded us of the fallibility of our spacecraft and engineers, may seem like an odd time to talk about heading off on some far more ambitious venture. Then again, it may be just the right time.

The shuttle is just a means to an end, rather than a destination in itself. The same goes for the space station. What excites the public, and what inspired the astronauts who gave their lives last week, is exploration. A space programme that doesn't dare to explore may ultimately be doomed to wither. ■





Hopes dashed
Space researchers
take stock after loss
of iconic mission
p562



Re-fusion
United States makes
move to return to
physics project
p563



Emergency cover
Harvard experts cook
up recipe to thwart
bioterror attacks
p564



Head's up
Wellcome Trust's
new chief to take
the reins in June
p567

Columbia explosion may trigger fatal delays for space station

Tony Reichhardt, Washington

The future of the International Space Station — the US\$25-billion project to place a permanent scientific laboratory in orbit — is hanging by a thread this week after the loss of the space shuttle Columbia cut its main lifeline to Earth.

If investigations into the accident (see below) drag on for more than a few months without a conclusive answer, or uncover pervasive problems that keep the remaining three shuttle vehicles on the ground, construction of the station will stall and the prospect that it will ever be fully crewed could permanently recede, experts on the project say.

On the other hand, if NASA can quickly locate and remedy the problem that caused the accident, the station could still be completed. And, either way, the overall impact of the tragedy on the agency's broader space-science programme may be small — in contrast to the retrenchment and disarray that followed the 1986 Challenger accident.

When Challenger exploded just seconds after launch, dozens of spacecraft, from the Hubble Space Telescope to the Galileo Jupiter probe, were stranded for years. But now, nearly all of NASA's astronomical and



The demise of Columbia has cast serious doubts over the viability of the space station programme.

planetary spacecraft — including the Space Infrared Telescope Facility and two Mars landers, scheduled for launch in April, May and June, respectively — go into space on expendable rockets instead.

That leaves the shuttle almost exclusively dedicated to building and servicing the space station, the US segment of which had been

scheduled for completion next February. The next shuttle flight, originally scheduled for March, would have delivered research equipment to the existing Destiny laboratory — including a freezer for storing specimens and a high-quality window for Earth observation. That was to have been followed by six more flights, spaced out over a year, to deliver

NASA sets up dual probes into shuttle accident

An internal NASA probe and an external committee of experts are already preparing the ground to find out exactly what caused Columbia to break apart over Texas.

The two teams will work in tandem with the help of hundreds of scientists and engineers to gather debris, analyse data and ultimately determine the cause of the crash. David Whittle, an administrator at NASA's Johnson Space Center in Texas, will head the internal probe, while the external investigation is being led by retired navy admiral Harold Gehman, who co-chaired the independent commission that looked into the terrorist attack on the USS Cole in 2000.

The swift decision to set up an external probe contrasts with the agency's handling of the 1986 Challenger disaster. NASA's silence in the first

few days after that accident led some to accuse the agency of withholding information.

This time around, NASA quickly convened not just an outside review, but frequent press briefings on the accident. "This is going to be the most open accident investigation people have experienced," pledges Michael Kostelnik, a senior administrator on the shuttle programme.

An early focus for the two investigations is the ceramic tiles on the shuttle's underside, which protect the heat-sensitive aluminium airframe during reentry into the atmosphere. Mark Lewis, a professor of aerospace engineering at the University of Maryland who specializes in re-entry physics, says: "I think it's conceivable that the loss of one tile in certain locations would be catastrophic."

Without ruling out other possible causes, investigators are looking at the possibility that the left wing of Columbia was damaged when debris from the shuttle's main fuel tank fell on it just after take-off. Programme manager Ron Dittmore says that a similar incident in 1992 left a several-inch-long gash in heat tiles on Columbia's wing, but it was concluded that the incident would not have endangered the shuttle.

Albert Wheelon, a physicist who served on the presidential commission that investigated the Challenger accident, predicts that investigators will soon move on from technical causes to possible management failures. "The first problem is to find out what went haywire," he says. "Then they must determine if there were hints of this beforehand, or if it was a bolt out of the blue."

Geoff Brumfiel

J. MITCHELL/REUTERS; WFAA-TV/APTN/AP (INSET)

▶ large pieces of the station's backbone-like structure and more solar arrays for electrical power.

That schedule has now been scrapped. But if the shuttles are flying again within a few months, the station's research programme may not be greatly affected, says Milburn Jessup, a professor of oncology at the Georgetown University Medical Center in Washington DC and chairman of the space station scientific users' committee. Only one of next year's assembly flights was to involve Columbia, and NASA might well be able to use another vehicle in its place.

Delays beyond that could pose problems, particularly if the sequence in which research equipment is delivered to the station has to be rearranged, says Jessup. A lengthy delay, like the almost three-year downtime after the Challenger accident, could be devastating, forcing scientists to stop work on experimental equipment bound for the station.

NASA is under pressure to get the shuttles flying quickly. The station's current three-man crew can remain onboard at least until June, their time being limited by factors such as onboard supplies. If the shuttle is not cleared to fly by then, NASA could, in a worst case, be forced to power down the outpost and leave it abandoned until a new shuttle crew can be sent up.

Russian Soyuz craft are currently the only vehicles besides the shuttles that can ferry crews to the station. But the cash-strapped Russian government is barely able to produce the two vehicles a year required to fulfil its obligations to the international project. And US law bars NASA from buying more Soyuz vehicles from Russia, which it accuses of selling arms to unfriendly states.

Watching the situation warily are the station's partners from Europe, Japan and Canada. Even if the shuttles resume flying soon, NASA still has no short-term plan for a vehicle that could safely evacuate a crew of six or seven — the minimum number required to make full use of the laboratory, according to scientists. NASA's newly proposed Orbital Space Plane, designed to dock with the station after launching on a conventional rocket, could take up to ten years to design and build.

Critics of NASA, such as Congressman Ralph Hall (Democrat, Texas), the senior Democrat on the House Committee on Science, argue that this would be far too little, too late, to salvage the utility of the space station. As long as the crew 'lifeboat' issue remains unresolved, they charge, the station will be next to useless as a scientific laboratory — whether the shuttles are flying again or not. ■

Science had rare leading role in ill-starred shuttle mission

David Adam

The mission that tragically ended over Texas on 1 February was one of a vanishing breed — it was the first flight of the space shuttle in three years that was dedicated entirely to scientific research.

Columbia was carrying more than 80 separate experiments across the spectrum of science, from tests on the effects of microgravity on protein-crystal growth and ant behaviour, to orbital observations of dust blown from the Sahara Desert, which may affect weather and climate in other regions.

The mission also investigated the effects of space flight on astronauts' health. Some tests probed the cells that make up the immune system, and others examined ways to stop astronauts developing kidney stones. One experiment simply sought ways to help astronauts sleep more soundly.

These tests, like most of the rest of the mission's scientific payload, were led by NASA scientists. But a portion of the mission's scientific capability was handed over to paying customers, including the European and Canadian space agencies, who were each using the 16-day flight to analyse the effects of anti-osteoporosis drugs on bone cells in microgravity. The US Air Force also had a stake in the flight, testing communications technology that would identify enemy threats using miniature satellites. These experiments were held in a special pressurized research module — designed and built by Spacehab, a company based in Webster, Texas — in the shuttle's cargo bay.

The Israeli Space Agency, which sponsored the dust experiment, thinks that about 80% of the data from that experiment and the agency's other projects were beamed to Earth before the shuttle's demise. Those carrying out other tests, including some involving live animals, have lost their studies and may have to wait years for another opportunity.

The mission also marked an effort by NASA to capture the imagination of a younger generation. A set of experiments based on a menagerie of wildlife including bees, spiders, silkworms and fish embryos was being keenly monitored by pupils at schools in the United States, Australia, China, Israel, Japan and Liechtenstein. Other students in Germany handled the ground-based control experiment for separate space research into aquatic ecosystems (see *Nature* 421, 307; 2003).

This geographical diversity was reflected in the mission's seven-strong crew. One, Ilan Ramon, was the first Israeli in space, and a second, Kalpana Chawla, hailed from India. The flight was Chawla's second trip on Columbia, having flown in 1997. Ramon, a former fighter



Columbia's crew addressed scientific issues ranging from astronaut health to climatology.

pilot who flew in Israel's 1981 mission to destroy a nuclear reactor being built in Iraq, attracted much of the pre-flight publicity, and ensured that security was beefed up.

Two of the other crew members had flown in space before. Rick Husband piloted the shuttle Discovery to the first docking with the International Space Station in 1999, and Michael Anderson was on Endeavour when it visited the Mir station in 1998. Of the astronauts making their maiden trips, two — David Brown and Laurel Clark — were physicians and the shuttle's pilot, William McCool, was a former US Navy test pilot.

Before the accident, questions had been asked about the scientific value of the mission's estimated US\$500-million price tag. Such conjecture will now be joined by more serious doubts over the human cost. Some researchers argue that unmanned probes can deliver better value; others insist that the challenge, daring and undoubted glamour of crewed missions bring less prosaic advantages.

Zev Levin at Tel Aviv University is one advocate of crewed research. As principal investigator of the Saharan dust experiment, Levin says that the flight captured the public imagination as an unmanned satellite never could. "In the past year I've spoken to students many times and you wouldn't believe the interest," he says. "Maybe some of them will go into science instead of becoming lawyers." ■

Additional reporting by Haim Watzman, Jerusalem.

K. RONSTROM/REUTERS

Fusion project back on track as US returns to the fold

Geoff Brumfiel, Princeton

The United States is to enter negotiations to rejoin ITER, the international magnetic-fusion experiment, sharply reviving expectations that the US\$5-billion project will be built.

US energy secretary Spencer Abraham announced the move on 30 January during a visit to the Princeton Plasma Physics Laboratory in New Jersey. The United States will join this year's negotiations to select a site for the project, he said, and is expected to contribute about a tenth of ITER's total costs. The project's current members — Japan, Russia, the European Union and Canada — are soon to be joined by China, which announced its intention to participate a few weeks ago (see *Nature* **421**, 306; 2003).

The US contribution to the project would amount to about \$50 million a year over ten years — less than the annual \$80 million it spent during the project's design phase.

President Bush confirmed the decision in a statement. "The results of ITER will advance the effort to produce clean, safe, renewable and commercially available fusion energy by the middle of this century," he said.

The US decision to rejoin the project was warmly welcomed by fusion researchers around the world as an important step towards making ITER — formerly known as the International Thermonuclear Experimental Reactor — a reality. "The level of US fusion physics is extremely high, and we are very happy that we'll be working together," says Yoshikazu Okumura, a fusion scientist at the Japan Atomic Energy Research Institute.

ITER was originally conceived as a prototype reactor that would sustain a fusion reaction in a plasma of hydrogen isotopes magnetically contained in a doughnut-shaped device called a tokamak. The United States, Japan and the European Union each spent several hundred million dollars to draw up blueprints for the machine. But concerns over its \$10-billion price tag, questions about its ability to reach design goals, and general budget constraints caused Congress to withdraw funding for the project, cancelling US involvement (see *Nature* **394**, 511; 1998).

The remaining partners pared down the original plans to a less ambitious \$5-billion design, and the Bush administration expressed an early interest in rejoining the project (see *Nature* **415**, 247–248; 2002). The revised design won the support of the US fusion community last summer, and a cost



Spencer Abraham reveals US plans to rejoin an international effort to build a tokamak larger than the DIII-D in San Diego (inset).

review by the Department of Energy completed late last year said that its price tag was realistic. "Our estimates and simulations give us confidence that ITER will work," says Raymond Orbach, head of the energy department's Office of Science, who will lead the US team at ITER's next round of negotiations in St Petersburg, Russia, on 18 February.

"I think current partners will be encouraged by US participation," says Robert Aymar, head of the ITER project, who is based in Garching, Germany. But Aymar says that he hopes the United States will consider paying more than 10% of the project's cost.

The planning process has been slowed in the past by political in-fighting among partner nations. "Hopefully the Americans will give the process a kick forward," says Geoff Cordey, a fusion researcher at the Joint European Torus facility near Oxford, UK.

The next round of negotiations will primarily be concerned with assessing the four proposed sites in Japan, France, Spain and Canada. Murray Stewart, who heads ITER Canada, is confident that, with its new partners, a site will be selected in a timely fashion. "The negotiations are proceeding very quickly," he says, adding that he thinks construction could begin by 2006.

▶ www.iter.org

Fading support ends Europe's dreams of neutron supremacy

Quirin Schiermeier, Munich

Plans for the European Spallation Source (ESS) — a state-of-the-art facility for neutron science — are near to collapse this week, after Germany and Britain indicated that they would not support its construction.

Germany's research minister Edelgard Bulmahn was expected to tell the country's cabinet this week that investment in the E1.4-billion (US\$1.5-billion) facility would put too much strain on Germany's science budget. And last month, says Dieter Richter, scientific director of the ESS project, Britain quietly revoked support for it at a research infrastructure meeting in Brussels.

The ESS, which was first proposed more than ten years ago, was to be the most powerful neutron source in the world. An estimated 5,000 European scientists use neutrons to observe the detailed molecular structure of everything from plastics to proteins.

But a leaked document from a meeting last month of the European Strategy Forum on Research Infrastructures states that "there is not sufficient support from among the Member States" for the realization of the ESS.

As a result of these developments, the ESS project team will be dissolved later this year, says Peter Tindemans, chairman of the ESS Council. "It makes no sense, under the current circumstances, to carry on with technology development," he says.

The decision is a crippling blow for an area of science in which Europe has long excelled. "Even US science advisers admit that Europe has global leadership in neutron research," says Richter. "Without the ESS we will soon lose it to Japan and the United States, which are both building and upgrading powerful neutron facilities."

The last hope for the project is that a scaled-down version might be built with financial support from one of the regions that applied to host the original facility.



Peter Tindemans: forced to halt neutron project.

A former airfield near Selby, UK, and a greenfield site in Sachsen-Anhalt, eastern Germany, had been selected as potential sites, and both regions may continue to push for the construction of a neutron source.

▶ www.ess-europe.de/ess_js/index.html

Fury at plan to split historic biology archive

Rex Dalton, San Diego

The auction of an historic archive of molecular-biology documents, planned for the fiftieth anniversary of the publication of the DNA double helix, is drawing fire from scientists.

Top researchers who sold papers to a private California archive on the understanding that they would be kept together in a single, accessible collection are angry that Christie's plans to sell the papers as separate lots. And a collector who gathered the works together is threatening legal action to block the sale.

Christie's is scheduling the sale of 56 lots from the Jeremy Norman Molecular Biology Archive for 25 April in New York, 50 years to the day after *Nature* published Crick and Watson's paper describing the double helix.

It values the documents at between \$2.2 million and \$3.3 million. The lots include documents from Aaron Klug, Max Perutz, Rosalind Franklin, Francis Crick and James Watson. One lot is a signed galley proof of Crick and Watson's *Nature* paper.

"It is an outrage," says Klug, who won the Nobel Prize in chemistry in 1982 for developing crystallographic electron microscopy. "We were given promises, assurances, that the whole collection would be kept together."



Aaron Klug: private archive sale is 'an outrage'.

Norman declined to be interviewed, saying only that the sale "will be handled appropriately". Christie's officials say that Norman has title to the documents, and that a bid for the entire collection would be willingly entertained.

But Al Seckel, a neuroscientist at the California Institute of Technology (Caltech) in Pasadena and an amateur collector who purchased the documents for Norman, was furious when he learned of the auction. "This impacts on my reputation," he says.

Beginning in the late 1990s, Seckel travelled to Europe to buy molecular-biology papers, which he sold to Norman, a book dealer in San Francisco. Both men said they

wanted to create an archive that transcends national borders, putting the history of DNA under one roof, probably at Caltech. For each purchase of papers and sale to Norman, Seckel says there are specific agreements that the documents will be kept together for scholarly work.

But Francis Wahlgren, head of Christie's books and manuscripts department, says that he is "not aware of any agreement". Norman no longer wanted to be responsible for the archive, he adds. In June 2001, *Nature* reported concern among archivists that the documents might someday be sold at auction, which at the time Norman denied would happen (see *Nature* 411, 732; 2001).

Norman's archive plans seem to have changed after his failed 2001 effort to acquire Crick's entire document collection. When the sale collapsed, the Wellcome Library in London stepped in to buy them for US\$2.8 million in November 2001.

Last week, Seckel tried unsuccessfully to persuade Christie's to stop the sale. He says he will go to court if necessary to block it.

David Pearson, the Wellcome's librarian, says: "Our concern is that very important papers may be disbursed in the private market." The Wellcome is considering its options, he adds.

Harvard team suggests route to better bioterror alerts

Jonathan Knight, San Francisco

A simple improvement in the way health data are monitored for signs of a bioterror attack could speed up the process and cut the number of false alarms, says a team of specialists at Harvard Medical School.

The findings, reported online this week (B. Y. Reis, M. Pagano and K. D. Mandl *Proc. Natl Acad. Sci USA* doi:10.1073/pnas.0335026100; 2003), are expected to influence the current US drive to improve early-warning systems for such attacks.

Public-health officials fear that biological attacks may not be recognized until it is too late to prevent casualties. Smallpox and anthrax, for example, start with 'flu-like' symptoms, and the first victims are likely just to be sent home to rest.

Biodefence researchers have been looking at everything from patterns of hospital visits to sales of cough syrup. Dozens of systems are now being field-tested by state and local health departments across the United States. In theory, a sudden outbreak of disease, whether natural or deliberate, will register as a spike in the data, alerting health officials.

But the chief difficulty is separating this from day-to-day variation. "There are a lot

of bumps and noise in public healthcare data," says Ben Reis, a biosurveillance specialist with Harvard Medical School at the Children's Hospital in Boston. To prevent false alarms, most systems set the alert threshold so high that they risk missing the first signs of a real outbreak.

The standard approach is to forecast the number of emergency cases that hospitals have to deal with one day at a time, based on historical data. Departures from the forecast send an alert to a regional epidemiologist for further investigation.

Reis designed his system to look at the data a week at a time. He reasoned that the wider window would make it easier to disregard blips that might otherwise register as false positives. It should also spot rising trends earlier than the standard software.

He tested the approach with emergency records from the Children's Hospital, which comprise the main complaint of every patient who checked in from 1992 to 2002 — a total of more than 500,000 visits. Because there were no real outbreaks, Reis added simulated ones calculated to look like small or large releases of anthrax or smallpox.

The week-long average was able to reveal

outbreaks that the one-day system missed, Reis and colleagues report, because its detection threshold could be set much lower without triggering false alarms.

Marc Overhage, an expert in healthcare informatics at the Regenstrief Institute in Indianapolis, says that false alarms need to be eliminated. Investigations into possible outbreaks are expensive, costing an average of \$50,000, he says. Too many could render a system worthless. "We shouldn't build all these surveillance networks until we know they work," he says. "Reis is the only one running through how to do it best."



Anthrax investigation: false alarms are costly.

J. KING-HOLMES/SPL

K. LAMBERT/AP

Science feels the pinch as Bush budget targets defence

Erika Check, Washington

American science is set to be squeezed next year, after President George Bush proposed a budget aimed at holding down total spending while shifting emphasis to military and homeland-security needs.

Bush's proposal for the 2004 fiscal year, which begins on 1 October, would see the federal government spend a record \$123 billion on research and development (R&D) — with more than half of the money going to the Department of Defense (see chart, right). Outside the Pentagon, R&D increases by about 4% — in line with the rest of the federal budget, but well short of the increases that science agencies have enjoyed in recent years.

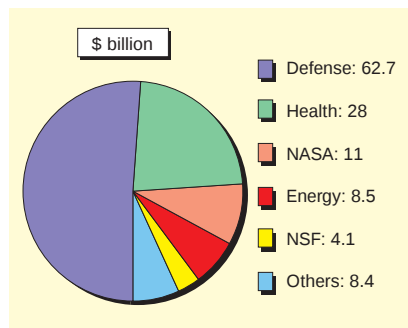
Announcing the research portion of the budget on 3 February, John Marburger, director of the White House Office of Science and Technology Policy, said that it reflects the administration's primary concerns of securing the nation against terrorist and military threats. "Science and technology are essential to the defence effort," he said.

Congress has not yet passed the 2003 budget, so it is difficult to compare Bush's figures with current spending. But his 2004 proposal for the basic research component of the R&D budget calls for a 5% increase in spending over his 2003 request.

If Congress, as expected, increases some agency budgets by more than Bush asked for last year, funding next year at most science agencies would be flat or increase only slightly. Most of the requested new spending will go towards security-related programmes.

"This budget raises some serious questions about the administration's overall commitment to science and technology," says Mike Lubell, director of public affairs for the American Physical Society. "At best, we're treading water."

The National Institutes of Health, which in recent years received double-digit increases,



Fighting fit: the Department of Defense is the winner in Bush's 2004 R&D budget proposal.

would get just a 2% increase over proposed funding for the current year. And although the president's request would double the number of biodefence-related grants to 660, it would add just 21 new grants to the 9,827 that will be available this year for research not related to biodefence.

"We have to ask to what extent the biodefence mission is altering public-health outcomes," says April Burke of Washington lobbying firm Lewis-Burke Associates. "The fear factor is what is being funded in place of research that will make Americans healthier."

At the National Science Foundation (NSF), which funds most non-biomedical research at US universities, the proposal favours fields that could be applied to intelligence and counter-terrorism — including mathematics, which is boosted by around 50% to \$89 million, and research into 'human and social dynamics', which would grow almost threefold to \$24 million. But the NSF's overall budget — although 9% more than what Bush proposed last year — would exceed the likely 2003 budget by just 2%.

The Bush proposal would increase the Department of Energy's Office of Science — which supports most US physicists — by just 1.4% over last year's request, to \$3.3 billion. High on the list of priorities for the science office is \$12 million in funding to support the international fusion experiment ITER, which the United States rejoined last week (see page 563). Computing, nanotechnology and microbial genomes would also do well.

When Congress finishes the 2003 budget, it will mull over Bush's proposal for 2004. Law-makers are due to send budget bills back to Bush to sign in autumn. In the past they have often given more money to science than presidents asked for, but that might not happen this year. "There's very little incentive for a Republican-controlled Congress to break much with this administration," says Burke. ■

Additional reporting by Geoff Brumfiel and Hannah Hoag.

Space research goes nuclear

An ambitious NASA budget proposal — crafted before the 1 February loss of the space shuttle Columbia — would make hefty investments in nuclear propulsion and high-energy astrophysics. Under the Bush proposal, NASA's budget would rise 3.1% to \$15.47 billion in 2004. The agency's nuclear-rocket initiative — now called Project Prometheus — would expand a programme that was restarted last year, spending \$279 million in 2004. This would fund a nuclear-propelled Jupiter Icy Moon Orbiter mission to search for subsurface oceans on the moons Europa, Ganymede and Callisto. The proposal also asks for \$59 million to set up a 'Beyond Einstein' astrophysics project, and \$39 million for developing fresh concepts and technologies for human space flight.

Ebbs and flows at standards lab

The Advanced Technology Program (ATP), a scheme to support early-stage industrial research and development, would end under the budget proposal. The Bush budget for the National Institute of Standards and Technology (NIST) would chop more than \$100 million from ATP's 2003 allocation, allotting just under \$30 million to wrap the programme up. But NIST would get \$70 million under the proposal to modernize and expand its main labs in Boulder, Colorado, and Gaithersburg, Maryland.

Falls for basic defence research

Despite massive expansion in the Department of Defense's efforts to develop new weapons systems, its spending on basic scientific research would be sharply curtailed under the budget proposal. Pentagon spending on basic research would fall by 8% to \$1.3 billion, and basic research spending at the Defense Advanced Research Projects Agency, which supports innovative and high-risk science, would drop by 25% to \$150 million. Meanwhile, the proposal expands the controversial ballistic missile defence programme by 17% to \$7.7 billion. The increase would allow development of new options for intercepting ballistic missiles and advanced radar systems to track incoming warheads, administration officials said.

Food safety comes to the fore

In a response to perceived threats from both animal diseases and terrorists, the Food and Drug Administration (FDA) and the US Department of Agriculture (USDA) would each get extra resources for food safety. The Bush proposal asks for \$20.5 million for food-safety programmes at the FDA, \$10.5 million of which would be used to start a registration system for food production, storage and shipping. The USDA also gets a 5.6% boost for food safety and \$47 million extra for lab security and research on animal diseases and vaccines.



Bush promises US cash for international battle against AIDS

Washington The United States should spend \$15 billion over the next five years to combat AIDS in southern Africa and the Caribbean, President George W. Bush said on 28 January in his State of the Union address.

The new scheme, called the Emergency Plan for AIDS Relief, would include money for drug treatments, prevention, testing and counselling in the two regions, which are home to 20 million people infected with

HIV. The funds would be distributed through networks of clinics. The plan will ultimately have to be approved by Congress, but seems to have support in both the Senate and the House of Representatives.

It is unclear what the announcement means for the future of the Global Fund to Fight AIDS, Tuberculosis and Malaria, which was launched in 2001 by the leaders of the world's eight most powerful nations. After making its second round of grants last month, the fund is nearly out of money. The United States has so far pledged \$500 million to the fund, and Bush's new plan includes \$1 billion in new donations through 2009. But critics say that this is a small sum compared with the \$6 billion that the fund says it needs over the next two years. Hopes that the United States may further boost its donations rose when Tommy Thompson, the US health secretary, was elected chairman of the fund on 31 January.

Report exposes errors in cancer 'breakthrough'

Munich Full details of the errors in an apparently successful cancer study have been revealed in a report from a committee at the University of Göttingen in Germany.

Problems with a trial led by Alexander Kugler, a former cancer researcher at

Göttingen (A. Kugler *et al. Nature Med.* 6, 332–336; 2000), were outlined by the committee last November (see *Nature* 420, 258; 2002), and the full report was leaked last month to the medical magazine *Deutsches Ärzteblatt*. Kugler's team claimed dramatic success with an experimental vaccine to treat kidney cancer, including remission in 7 out of 17 patients. But the report says that the paper's authors were unable to document the remissions satisfactorily, or had failed to mention the use of other relevant therapies. In one patient, for example, the authors claimed a "complete response", when the secondary tumour had actually been surgically removed after vaccination.

The DFG, Germany's main research funding agency, is investigating the case further to determine the responsibility that each of the paper's 15 co-authors and the department head should bear.

Student cries foul over evolutionary snub

Washington Do academics have the right to deny able students a letter of recommendation? The question is currently being debated at Texas Tech University in Lubbock, after student Micah Spradling, an aspiring physician, alleged that a biology lecturer denied him a letter because he did not believe in evolution.



Holding out for help: international pledges are a lifeline for South Africans infected with HIV.

Spradling has filed a complaint with the US Department of Justice, which has asked the university to respond to the allegations. The lecturer — Michael Dini, who teaches life sciences — has refused to comment. But Dini's website outlines the criteria that students must meet to obtain a letter of recommendation. Among them, students must "truthfully and forthrightly affirm a scientific answer" to the question, "How do you think the human species originated?"

Dini's critics accuse him of demanding that students believe in evolution. But others point out that his criteria only require students to answer a question about evolution in a scientific manner, and may not necessarily prevent a student who holds different beliefs from obtaining a letter of recommendation.

London medic makes short trip to head Wellcome Trust

London The Wellcome Trust, the world's largest biomedical charity, has named its new head.

Mark Walport will take over as director of the London-based institution in June this year. Walport is currently head of medicine at Imperial College, London, and has been a governor of the trust since 2000. He replaces Mike Dexter, who will leave the trust next month after five years in the position.

"The Wellcome Trust is a very special



Wellcome development:
Mark Walport.

£400 million in grants and support each year.

organization," Walport says. "Not only does it make huge contributions to biomedical research in Britain and internationally, it also plays an important role in supporting the medical humanities." The trust uses its £12-billion (US\$20-billion) endowment to support some 3,000 biomedical researchers, dishing out

European scientists plan jamboree in 2004

London Europe is to get its own biennial science knees-up, described by the organizers as a counterpart to the annual American Association for the Advancement of Science meetings.

The first such meeting, called the EuroScience Open Forum, will take place in Stockholm in August 2004. It will feature lectures, interviews and debates aimed at promoting cutting-edge science and stimulating debate on the role of science in society. EuroScience, a Strasbourg-based organization that promotes science, is also promising a science film festival and exhibitions on science and the arts.

The organizers hope to attract around 200 speakers and are inviting interested participants to submit ideas for events.

► www.euroscience.org

Nobel winners rally against unilateral attack on Iraq

Washington More than 40 Nobel prizewinners from the United States have voiced objection to their country going to war with Iraq unless it first secures broad international support. The Nobel laureates have published a declaration on the Internet denouncing such an attack, and are inviting individuals and organizations to register their support.

The protest was organized by Walter Kohn of the University of California, Santa Barbara, a former defence adviser to the Pentagon and winner of the 1998 Nobel Prize in Chemistry. Other signatories include 1989 physics winner Norman Ramsey, a former Manhattan Project scientist, and Harold Varmus, former director of the National Institutes of Health who won the 1989 medicine prize.

The statement points out that "war is characterized by surprise, human loss and unintended consequences". Even if an attack is successful, it states that this "would undermine, not protect, US security and standing in the world".

► www.nobel laureatesoniraq.org

Killing fields: poor farmers need new crop strains that can tolerate extremes such as drought. But the valuable skills of plant breeders (inset) are under threat.

E. FELIX/AP

A dying breed

Public-sector research into classical crop breeding is withering, supplanted by 'sexier' high-tech methods. But without breeders' expertise, molecular-genetic approaches might never bear fruit. Jonathan Knight reports.

Normally, at this time of year, agricultural scientists from around the world would be converging on the headquarters of the International Maize and Wheat Improvement Center, known as CIMMYT, in Texcoco, near Mexico City. They would then travel together to a desert field station near Ciudad Obregón in northwestern Mexico to study the current crop of experimental wheat cultivars, planted at the beginning of winter.

But not this year. For the first time in half a century, the research centre that helped to sow the seeds of the 'green revolution' of the 1960s and '70s has been forced to skip a cycle of wheat breeding trials, because of a lack of money. More than half of CIMMYT's fields in Obregón lie fallow, and the trainee plant breeders are staying at home.

CIMMYT is not alone. All over the world, conventional plant breeding has fallen on hard times, and is seen as the unfashionable older cousin of genetic engineering. "Plant breeding is getting dumped along the wayside for not being sexy enough," claims Greg Traxler, an agricultural economist at Auburn University in Alabama. Government funding of plant-breeding research has all but dried up in the United States and Europe, and the World Bank and donor nations have recently slashed their support for the Consultative

Group on International Agricultural Research (CGIAR), the international research consortium of which CIMMYT is a part.

Meanwhile, a steady push by companies to claim exclusive commercial rights to new plant varieties has progressively tied the hands of publicly funded efforts at crop improvement. If this trend isn't halted, some experts claim, tomorrow's supercrops may end up like many of today's medicines: priced out of the reach of much of the developing world's growing population. "We are headed down the same path that public-sector vaccine and drug research went down a couple of decades ago," says Gary Toenniessen, director of food security at the Rockefeller Foundation in New York.

Sowing success

Classical breeders improve crops simply by crossing plants with desired traits, and selecting the best offspring over multiple generations. Sometimes they use chemical mutagens to disrupt crop genomes, in the hope that some of the resulting mutants will have useful new traits. Crosses may be as simple as letting two plants grow together, or they may require pollination by hand. And for crops such as wheat, one parent must first be emasculated to prevent self-pollination. In some ways, breeding is like accelerated,

targeted evolution, and as long as test crops and seed banks are maintained, the possibilities can never be fully exhausted.

These methods, applied intensively at CIMMYT and the International Rice Research Institute (IRRI) near Manila in the Philippines, provided the impetus for the green revolution. Breeders produced dwarf varieties of wheat, maize and rice that were less likely to fall over in wind and rain, and which could carry larger seeds. Thanks to these varieties, farmers could use more fertilizer without risking losing their crops, and grain harvests in some areas have doubled or even trebled over the past three decades.

Central to CIMMYT's success in wheat was the practice of 'shuttle breeding', in which two seasons of plant selection could be completed in one year. Grain would be rushed from the fields in Ciudad Obregón after the harvest in April for summer planting in Toluca, near Mexico City.

This year's cancellation of the Obregón end of the shuttle was part of a 10% reduction in CIMMYT's programmes in the face of budget cuts, says the centre's director general, Masa Iwanaga. This was a result of the reduction in support for the CGIAR, which supports CIMMYT, IRRI and 14 other agricultural research centres around the world.

Whereas the CGIAR's funding crisis has

come to a head in the past couple of years, exacerbated by the global economic downturn, the world's academic plant-breeding labs have suffered steady attrition over a far longer period. Molecular genetics and transgenic technologies hold great promise for crop improvement, and have consumed a growing portion of the limited funding pie. University administrators have reinforced this trend, tending to replace retiring plant breeders with molecular geneticists who are more likely to produce high-profile journal articles.

Changes in the intellectual-property environment have also taken their toll. From the late 1960s onwards, developed nations introduced a legal framework of plant breeders' rights, giving new varieties and cultivars patent-like protection. The goal was to stimulate innovation in corporate labs, but the reforms also meant that public-sector breeders were no longer free to tinker with plants grown from commercial seed. "Plant-variety protection was the death knell for public breeding programmes," says Michael Gale, head of comparative genetics at the John Innes Centre in Norwich, Britain's leading public plant-science research institute.

Root of the problem

The figures reinforce Gale's view: until the 1960s, breeding for crop improvement was largely a public endeavour, but a survey of US plant scientists in the mid-1990s found more than twice as many breeders in the commercial sector than at universities and government agencies combined¹. And although breeders' skills are still alive in the private sector, they are now working to subtly different ends. For seed companies and agribiotech firms, the top priority has been developing crops that can maximize profits from the intensive agricultural practices that are widely used in the developed world. Sadly, there is less money to be made in seeding a second green revolution for the world's poor.

In recent years, of course, the big news in the commercial and public sectors has been transgenic technology, rather than conventional breeding. Genetically modified (GM) crops that are resistant to the effects of broad-spectrum herbicides or that carry genes for insecticidal toxins have been widely planted across North America — but simultaneously shunned by European consumers, who are deeply suspicious of the technology. The welter of media coverage has obscured recent achievements in classical breeding, and although breeders generally view transgenics as a valuable tool, they stress that conventional breeding is far from obsolete.

In fact, for many GM crops, there is a comparable conventionally bred variety. The seed company Pioneer Hi-Bred, based in Des Moines, Iowa, for instance, produces a conventional, herbicide-resistant oilseed rape, or canola, that has similar advantages for weed control as its GM counterparts. And whereas

the GM 'golden rice'², engineered to contain a gene that boosts the production of vitamin A by people who eat its grain, has attracted much publicity, conventional breeding is also being deployed to improve the nutritional value of this staple crop. IRRI has produced a cultivar of rice called IR68144 that bears grain rich in iron³, and so could be used to combat anaemia. Even for crops such as the banana, which is unable to reproduce sexually without specialist human intervention, conventional breeding may still have a role to play (see "Bananas in the fertility clinic", below).

What's more, the GM crops developed so far generally involve only the addition of a single gene. Looking to the future, it's unclear whether complex traits, which are thought to involve multiple genes, will be amenable to manipulation through genetic engineering. "In the long term, you need heat tolerance, salt tolerance, greater yield and so on," says Paul Gepts, a crop geneticist at the University of California, Davis. "Some say you can do it with genetic engineering, but we just don't know how those systems work and how those genes interact." By contrast, practical experience has shown that conventional breeding can be used to improve a suite of subtle traits simultaneously.

All of this makes Donald Duvick, who was head of research at Pioneer Hi-Bred until

his retirement in 1990, concerned about the future of crop improvement should the agribiotech giants lose their enthusiasm for transgenics. "I worry that the results will be so far in the future that industry will say 'we can't wait that long,'" he says. If so, the depleted public-sector effort in plant breeding may be ill-equipped to take up the slack.

There are already hints that some companies are pulling back from long-term investments in high-tech crop improvement. Only last month, the Swiss-based multinational Syngenta closed its Torrey Mesa Research Institute near San Diego, which was a major force in crop genomics. And both Syngenta and its US rival DuPont, which owns Pioneer Hi-Bred, have recently withdrawn funding from the John Innes Centre. "The industry is in turmoil," says Gale.

Against this sombre background, can anything be done to safeguard future progress in crop improvement by reviving the science of plant breeding in the public sector? There is no easy answer, but some experts suggest that the future lies in boosting the power of conventional breeding by marrying it to genomic and other molecular-genetic techniques, while making a concerted effort to break with the proprietary approach to intellectual property that is currently blighting the field.

Bananas in the fertility clinic

Having shunned sex for thousands of years, bananas are in trouble. Those grown commercially are sterile mutants, propagated by replanting the suckers that sprout from existing trees. Lacking the genetic shuffling of sex, the single variety that dominates the export market is susceptible to any pest that evolves to evade its defences against disease.

In the late 1990s, the emergence in Southeast Asia of a new strain of Panama disease, a wilt caused by the fungus *Fusarium oxysporum*, devastated commercial plantations. It has since spread to Australia and Africa, and if it lands in Latin America, where most export bananas are grown, farmers will need a new resistant variety.

Genetic manipulation seems the obvious answer — and researchers at the Catholic University of Leuven in Belgium have already produced several transgenic varieties that carry genes for antifungal proteins⁵. These will be field-tested for resistance to Panama disease over the next few years.

But even if they pass these tests, there is no guarantee that Europe's suspicious consumers will warm to the idea of transgenic bananas. So a conventional breeding effort is also under way. Breeders at the Honduran Agricultural Research Foundation in San Pedro Sula have found that it is just about possible to breed bananas, through careful hand pollination and sieving hundreds of tonnes of banana pulp to collect the few resulting seeds.



One beacon of hope comes from a consortium of researchers at 12 institutions headed by Jorge Dubcovsky, a wheat molecular geneticist at the University of California, Davis. Its primary tool is 'marker assisted selection' (MAS). This technique, enthusiasts claim, could offer to plant breeding what the jet engine has brought to air travel. Traditionally, breeders have relied on visible traits to select improved varieties. For pest resistance, for example, that means examining mature plants in the field over successive generations to see which survive best in the face of attack by pests, before carrying out new crosses. MAS, however, relies on identifying marker DNA sequences that are inherited alongside a desired trait during the first few generations. Thereafter, plants that carry the trait can be picked out quickly by looking for the marker sequences, allowing multiple rounds of breeding to be run in quick succession.



Jorge Dubcovsky's genetic techniques aim to give traditional breeding a technological boost.

Superior breeding

MASwheat, as the consortium is known, aims to select for 23 separate traits in wheat, conferring resistance to fungi, viruses and insect pests. Its members also hope to breed the grain to produce bread and pasta of superior quality. Notably, the consortium is making all of its marker sequences and research protocols freely available. "If you go to our website, you have all the tools to do this anywhere in the world," Dubcovsky says.

For wheat, this admirably open approach was relatively easy to adopt, because it is one of the few crops to remain largely in public hands. Because wheat is self-pollinating, many farmers simply plant a portion of their harvest each year, safe in the knowledge that it will retain its desirable characteristics. Not surprisingly, this has restricted the interest of commercial seed producers, who don't see a robust market for their products.

Elsewhere, however, intellectual property is creating a heavy burden, with universities

and other institutions facing barriers to the free exchange of seed, and restricted access to cutting-edge molecular technologies. "I wish it would all go away," says Kent McKenzie, director of the California Rice Experiment Station, which develops new varieties of the crop in its test fields at Biggs, north of Sacramento.

Extending the MASwheat consortium's approach to other crops may require public institutions to band together to end the practice of granting exclusive licences to individual companies each time they develop a powerful technology for crop improvement. To this end, Toenniessen has been meeting with representatives of ten 'land grant' universities — which form the backbone of agricultural research in the United States — to hammer out a plan. "If those in the public sector worked collectively, they could solve their problems," says Toenniessen. He hopes to pioneer the approach in speciality crops such as peanuts, broccoli, lettuce and tomatoes, in which the seed and agribiotech industry does

not have strong commercial interests.

Richard Jefferson would go further. His Center for the Application of Molecular Biology to International Agriculture (CAMBIA) in Canberra, Australia, is trying to put cutting-edge technology for crop improvement directly in the hands of developing-world scientists and farmers, rather than leaving them to depend on the continued health of labs in rich countries. "The money is drying up and that is not going to change," he says. "We need to rethink the way crop improvement is done."

In part, Jefferson says, this will involve the transfer of transgenic technologies. But extending access to molecular-genetic enhancements to conventional breeding methods will also be crucial. Researchers at CAMBIA, for instance, have developed a DNA microarray that will boost MAS. In many crops, it is difficult to search for specific genetic markers, because very little of their DNA has actually been sequenced. But by immobilizing fragments of DNA from a variety of cultivars on a microarray and then seeing which of them bind to DNA sampled from individual plants, it is possible to look for the presence of genetic markers in these plants in the absence of any sequence information⁴.

This technology has already been adopted by the International Center for Tropical Agriculture in Cali, Colombia, for cassava improvement. "It is extremely useful," says Joe Tohme, the centre's director of biotechnology. By making such techniques freely available, and allowing scientists anywhere in the world to tinker with and improve them at will, Jefferson hopes to speed progress. Essentially, he wants to create a crop-improvement counterpart to the 'open-source' software movement that has managed to flourish alongside the proprietary approach of giants such as Microsoft, which keep their programs' codes under wraps.

'Open-source molecular agronomy' is certainly a sexier label than conventional plant breeding. But will it have sufficient cachet to reverse the current decline in public-sector crop improvement? The food supply for future generations in the developing world could hinge on the answer. ■

Jonathan Knight writes for *Nature* from San Francisco.

1. Frey, K. J. *National Plant Breeding Study* (Iowa Agric. Home Econ. Exp. Station, Ames, Iowa, 1996).
2. Ye, X. *et al. Science* **287**, 303–305 (2000).
3. Glahn, R. P., Chen, S. Q., Welch, R. M. & Gregorio, G. B. *J. Agric. Food Chem.* **50**, 3586–3591 (2002).
4. Jaccoud, D., Peng, K., Feinstein, D. & Kilian, A. *Nucl. Acids Res.* **29**, 25e (2001).
5. Remy, S., Francois, I., Cammue, B., Swennen, R. & Sági, L. *Acta Horticult.* **461**, 361–365 (1998).

CIMMYT

♦ www.cimmyt.cgiar.org

IRRI

♦ www.irri.org

MASwheat

♦ maswheat.ucdavis.edu

CAMBIA

♦ www.cambia.org



Free for all: Richard Jefferson wants to put crop improvement within the reach of poor farmers.

Out of the kitchen

Designer microwave ovens that can heat reactants in record time are heralding a quiet revolution in chemical synthesis. David Adam feels the heat.

They may still struggle to turn out a decent baked potato, but the twin virtues of speed and convenience have ensured that there is barely a modern kitchen without a microwave oven. And the same could soon be true of chemistry laboratories. A new generation of devices designed specifically for chemical synthesis is driving fresh interest in the power of microwaves among industrial and academic chemists alike.

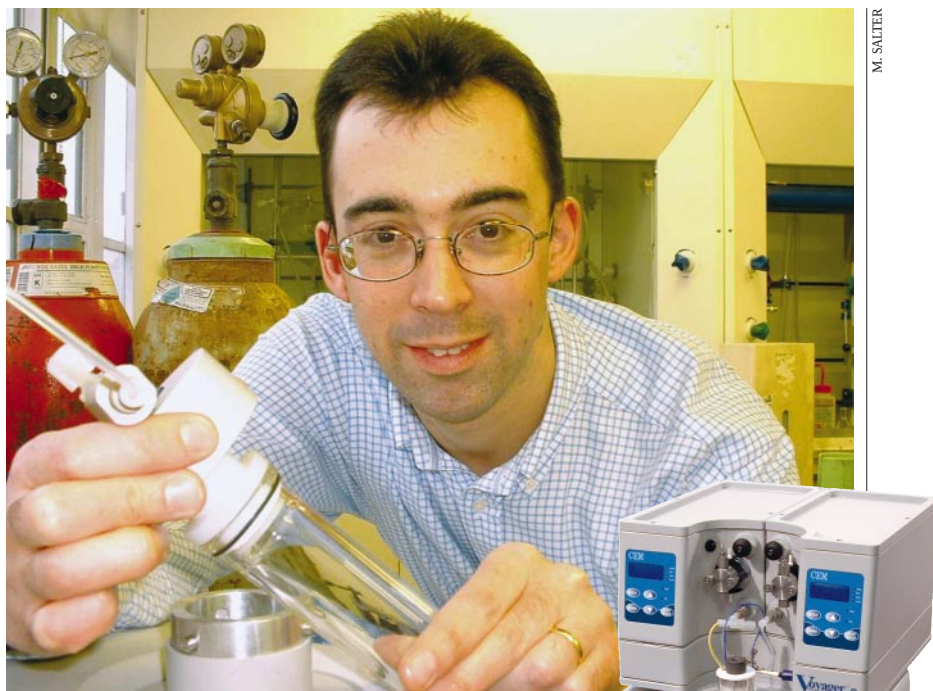
As in the kitchen, speed is the strongest selling point for the devices. Microwaves heat reactants much more quickly than conventional means, so syntheses that usually take hours can be done in just a few minutes. Other reactions can be made more selective, producing more of a desired molecule and less contaminating by-products. "In 10–15

years, we will see a microwave reactor in every academic and industrial laboratory," predicts Oliver Kappe, who works on microwave synthesis at Graz University in Austria.

"They will be the Bunsen burners of the twenty-first century."

Microwaves have been used sporadically in scientific research

Hotting up: Oliver Kappe believes the demand for microwave devices in chemistry will grow.



What's cooking?
Nicholas Leadbeater uses the new microwave ovens (inset) to revamp reactions.

since the first commercial ovens were introduced in the 1950s. But it was not until the 1980s that chemists began to get in on the action. Given the precision usually associated with chemical synthesis, the techniques involved in these early experiments were remarkably crude. "You put your reactants in a sealed vessel inside a conventional domestic oven, bedded it down in sand and blasted it with 'defrost', 'cook' or whatever," recalls Nicholas Leadbeater, an organic chemist at King's College London.

Accidents were common. "Everyone who has done chemistry in a domestic microwave oven has had some kind of mishap," Leadbeater says. "Depending on what you were doing, either the vessel exploded inside the oven cavity or you just blew the door off." Those chemists who braved the risks saw interesting results — usually a phenomenal increase in reaction rate. But with no way to measure the temperature and pressure inside the oven, and frustrating variability between the results obtained using different models, experimental findings could rarely be reproduced reliably.

Hot, hot, hot

Things began to get more precise two or three years ago, when companies making specialist microwave equipment for applications in the food industry — such as breaking down fats for food-quality testing — realized that synthetic chemists would pay good money for more reliable machines. Chemists working in the pharmaceutical industry were the first to benefit, but since then a growing number of academic labs — including those of Kappe and Leadbeater — have also invested

in the new tabletop devices.

The ovens allow accurate control of temperature and pressure, and include design features to contain exploding reactants — "we still have about one explosion a week," Leadbeater admits. With these microwave reactors, synthetic chemists are now revisiting notoriously slow or temperamental reactions.

Leadbeater's group, for example, has reworked an important synthesis called the Suzuki reaction, which forges a link between two benzene rings to make 'biaryl' molecules. These organic structures are found in a wide range of compounds including pharmaceuticals, polymers and liquid crystals. Shifting the synthesis into a microwave reactor removes the need for an expensive palladium catalyst, making the process cheaper and less prone to contamination with traces of palladium¹.

Microwave heating can also steer reactions towards desired products. Using conventional heat sources, for instance, the reaction between naphthalene and sulphuric acid produces an equal mixture of two possible isomeric forms of a molecule that is used in the manufacture of dyes and other industrially important chemicals. But the increased rate of heating obtained using microwaves means that the reaction can be driven to give one or other of the products². Other researchers are deploying microwaves in the name of 'green' chemistry, performing

syntheses that usually require toxic or flammable solvents in just water, or even with no solvent at all³.

But in most cases it still takes a great deal of trial and error to optimize the conditions for microwave synthesis. "You have to put in the time to get the conditions right," Leadbeater says. "Too much power and you absolutely obliterate your materials; too little and nothing happens. But once you get it right, you're away." Indeed, after a set-up has been perfected, dozens of reactions can be run one after the other. The ability to halt reactions before they have the chance to produce contaminating by-products can also be important, and the new microwave reactors allow chemists to do this by cooling the tube containing the reactants with a blast of cold air.

Microwaves also have the useful property of heating some materials while leaving others cold — as anyone who has eaten a jam tart heated on full power for a minute in a domestic oven will testify. Essentially a high-frequency electric field, microwaves cause free ions or electrons to move in the same direction as the applied field. For polar molecules, such as water, which have an uneven distribution of electrical charge, the applied field forces them to wobble. Both of these types of movements generate heat. But other materials don't respond strongly to the field, and remain cool.

Inorganic chemists have taken advantage of this to make compounds such as metal sulphides, which have a variety of applications including as catalysts and metal coatings. Conventionally, the powdered metal and sulphur are simply mixed and heated in a sealed tube. But because the sulphur vaporizes and quickly increases the pressure, explosions are common — which means that the reaction must be carried

The advantage of speed means that microwave chemistry is set to be a growth area.

out cautiously over several days. Microwaves heat only the metal, which then conducts the heat to the sulphur, allowing a less riskier reaction to take place. And — ping — the chemists get their metal sulphide in just 15 minutes (ref. 4).

Using conventional heat sources, the energy must first be conducted through the walls of the vessel containing the reactants. But microwaves heat the contents directly, allowing the temperature to rise much faster, and so boosting reaction rates. In fact, in many cases the increase in reaction rate is so great that researchers initially speculated that there was a strange effect at play, unrelated to the rapid supply of heat. This apparent 'microwave effect' puzzled chemists, some of whom invoked unconventional ideas — such as microwaves causing a reduction in the amount of energy needed to start a reaction — to explain it.

Supercharged

Most experts now believe that the mysterious effect is the result of nothing more than microwaves' ability to superheat solvents way beyond their normal boiling points — because the even spread of heat through the liquid allows it to reach a higher temperature before bubbles form. Water, for example, reaches 105 °C before boiling in a microwave oven; whereas the solvent acetonitrile boils at 120 °C instead of its usual 82 °C. "There are

no general non-thermal microwave effects," asserts Kappe. "There is no magic."

Still, some scientists believe that questions about microwave mechanisms remain. Gavin Whittaker, a chemist at the University of Edinburgh, UK, who uses microwaves to prepare solid-state materials, is one. In unpublished work, he has observed "anomalies" in the microwave synthesis of barium titanate, used to make capacitors in the electronics industry. In microwave reactors, he says, the rate of reaction in many syntheses is so high it cannot be accounted for by the heating effects alone. "We've got evidence that there's an enhancement of diffusion of ions in the direction of the electric field," he claims.

Whittaker has presented his results at conferences, where some sceptics have suggested that they are an artefact caused by problems with the temperate measurements — which are notoriously difficult in solid-state systems. But Whittaker insists that the phenomenon is genuine. "It needs to be examined rigorously and properly, rather than people just saying it's down to bad measurements," he says. Other researchers have noted similar effects, most notably in the preparation of compounds such as aluminium oxide from powdered ingredients. John Booske and Reid Cooper, materials scientists at the University of Wisconsin-Madison, for example, argue that microwave heating increases the driving force for ionic diffusion when powdered ingredients are processed into ceramic materials⁵.

Whittaker believes that research into the mechanisms of microwave action is the most intriguing aspect of the field. "The exciting work at the moment is in understanding the fundamental principle of what happens," he argues. But even if such effects are found to be artefacts, and there are no new physical phenomena to explore, the simple advantage of speed means that microwave chemistry is poised to become a growth area.

"Industry has really caught on, but what's holding the academic field back now are the prices," says Kappe. The cheapest of the new generation of microwave reactors currently sells for about US\$20,000, beyond the buying power of many small academic labs. But Kappe is confident that market forces will soon rectify that situation. "As more and more academic chemists hear from their colleagues in industry about these devices, they will want to have one," he says. "And then competition between equipment suppliers will drive the prices down."

David Adam is a news and features writer for Nature.

1. Leadbeater, N. & Marco, M. *Angew. Chem. Int. Ed.* (in the press).
2. Stuerger, D., Gonon, K. & Lallemand, M. *Tetrahedron* **49**, 6229–6234 (1993).
3. Varma, R. S. *Pure Appl. Chem.* **73**, 193–198 (2001).
4. Whittaker, A. G. & Mingos, D. M. P. *J. Chem. Soc. Dalton Trans.* 2073–2079 (1995).
5. Booske, J. H., Cooper, R. F., Freeman, S. A., Rybakov, K. I. & Semenov, V. E. *Phys. Plasmas* **5**, 1664–1670 (1998).

Gavin Whittaker is keen to discover the mechanisms behind the rapid reaction rates obtained using microwave heating.

Time for bioethics and business to start talking

Commerce and idealism need not be mutually exclusive.

Sir — The cultural rift between bioethicists and biotechnologists has resulted in a situation where each sees the other as single-minded. It is all too easy for bioethicists to assume that businesses are entirely driven by the profit motive, with science merely a slave to this machine. For their part, many corporations deliberately exclude themselves from debating ethics, claiming that their expertise is grounded in the reality of (say) healthcare needs or food production, and dismissing bioethicists as caught up in moral abstractions.

Both groups need to realize the imperative for interaction, particularly if they are truly dedicated to social responsibility. Perhaps the best way to achieve this is to explain how each field can better serve its respective goals by building a constructive rapport. Regrettably, there are only a few on each side of the fence who have come to this realization. Each group must overcome its misgivings, embrace common goals, and try to appeal to the sensibilities of the other.

Put simply, bioethicists and business people need to realize how valuable they can be to each other while honouring their own professions.

Industry representatives are well aware that to meet their goals of stakeholder satisfaction, sustainability, profit and continuing research, they have to understand the intricacies of the life-science landscape. This requires understanding patient interests, legal and regulatory trends, market segmentation and international receptiveness to products. In the most general sense, bioethics offers a context for technology — socially, economically and politically — which is critical given the complexities of the biotechnological enterprise.

If bioethics offers industry a context, how can industry reciprocate? The answer is that it offers relevance to bioethics. Bioethics, amongst other things, is concerned with the just development and distribution of medical and other life-science technologies. The discipline has

become concerned with setting policies for scientific grant funding, academic research, and private and public hospitals. But when asked to address the bioscience industry, the question becomes eclipsed by debates over the ethical propriety of working with private firms rather than how to engage them. For instance, helping corporations to understand the ethical context of their research is seemingly less important than holding endless discussions over conflicts of interest.

I think it is important for bioethicists to understand the part of the object of its study that is substantially influenced by commercial factors. Failure to engage industry will eventually lead the discipline towards irrelevance, as it constructs artificial boundaries based on fashions that neglect the important and pervasive role of commerce.

Rahul K. Dhanda

*Bioethics Program, Interleukin Genetics,
135 Beaver Street, Waltham,
Massachusetts 02452, USA*

Tools for modelling biological processes

Sir — Without minimizing problems presented by large-scale ranges, as indicated in H. Kitano's good overview article in your Computational Systems Biology Insight (*Nature* 420, 206–210; 2002), I think that the potential leverage offered by a 'layered platform' approach to modelling biological processes deserves some mention.

Members of the Biomedical Information Science and Technology Initiative (BISTI) of the US National Institutes of Health (NIH) concluded in 1999 that if we properly develop software tools, "time spent reinventing the same [software] processes in one laboratory after another will be freed for basic research" (see www.nih.gov/about/director/060399.htm). The NIH subsequently started a programme focusing on tool development, called Innovations in Biomedical Information Science and Technology, and, more recently, established the National Institute of Biomedical Imaging and Bioengineering.

How exactly does one avoid reinventing the wheel? As an illustration, transport phenomena are involved in biological/physiological processes whose scale ranges from individual protein production, via tissue nourishment and organ functionality, to whole-organism homeostasis. While

transport is the common underlying theme, each system differs in geometry and the properties of materials involved.

A single 'transport platform' that allows a user to configure the geometry with finite element analysis and to specify 'material' behaviour with specific constitutive equations could simulate applications in all biological transport systems. The remaining 'layers' of the total solution — fluid mechanics, chemical reactions, heat transfer, matrix inversions and so on — would be supplied by experts in these disciplines and managed automatically.

The feasibility of this approach, and its compelling economic logic, is supported by the success of analogous platforms in other venues, for example ANSYS in structural mechanics (see www.ansys.com). What is perhaps not so obvious is that biologists, physiologists, clinicians and others can model phenomena that they understand without being required to make a major time investment in such matters as analytical mechanics, optimum matrix inversion techniques, database management or graphical user interfaces.

As we begin the effort to develop computationally intensive models of biological phenomena, it is worthwhile to pause for a moment and think about the structure of the tools that are needed.

John Keane

*JKA Inc, 273 Jefferson Road, Princeton,
New Jersey 08540, USA*

Complexity may teach us a simple lesson

Sir — In his Concepts essay "Progressive evolution: aspirational thinking" (*Nature* 240, 611; 2002), Henry Gee failed to mention the most obvious source (perhaps the second most obvious, after our own vanity) of the idea that evolution is progressive. This is that evolution has actually been progressive, insofar as it has produced increasingly complex organisms over time.

It may be that this has occurred by way of undirected processes, but one can hardly blame people, including scientists, for arriving at the conclusion that evolution is progressive, given its history here on Earth. Of course, this requires a value judgement that more complexity is better, but it is not surprising that human beings will think that the complexity of a human is 'better' in some sense than that of a cockroach.

Michael Sierk

*Department of Biochemistry and Molecular Genetics, University of Virginia, PO Box 800733,
Charlottesville, Virginia 22908-0733, USA*

correspondence

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited for length.

Fuel for thought

Stanford has signed up for a ride on a tiger. Will it survive the journey intact?

David Ritson

On 20 November 2002, Stanford University announced a \$225-million, ten-year sponsorship deal with four giants of the energy industry — ExxonMobil (Esso), General Electric (GE), Schlumberger and E.ON, a major German energy-supply company. The result is the Global Climate and Energy Project, G-CEP. The amount of money is equal to the total of all industrial grants to Stanford over the past ten years, and annually is an order of magnitude larger than each of the current industrially sponsored climate-projects at Massachusetts Institute of Technology and Princeton.

The G-CEP is billed by Stanford as an attack on “one of the grand challenges of this century”, namely the problem of how to increase world energy supplies “while substantially decreasing greenhouse emissions”. Who better to do this than Stanford? On the surface, this is a win-win situation, with Stanford getting substantial research money and the sponsoring companies refurbishing their tarnished and battered environmental images.

If this was all that was at stake, the deal would be great. After all, it is not the role of a university to pass judgment on the character of its donors. But in this instance, Stanford University has shown a marked blindness to the political agenda of the sponsors, which is a commitment to foster and expand the use of fossil-fuel energy (over the next ten years, \$100 billion is being spent on oil exploration by ExxonMobil alone), while holding out the hope that technology will solve the resultant greenhouse-gas emission problems. If Stanford had faced up to this agenda, would it have accepted the proffered funds? No, except under conditions guaranteeing independence from the sponsors’ political agenda and accepting the \$225 million in the form of an unrestricted gift to a climate-change programme. Unfortunately, the G-CEP was not set up in this way.

History of the G-CEP

The G-CEP originated a year and a half ago when Franklin Orr, a professor of petroleum engineering and at that time dean of Stanford’s Earth sciences department, suggested a research collaboration on carbon sequestration to Schlumberger, a large oil-service company. The objective of physical carbon sequestration is to inject CO₂ produced from fossil fuels into deep geological aquifers or the ocean depths, instead of venting it into the atmosphere. This technique has been



Greenhouse gases: does the solution for our environment lie in prevention or cure?

used since 1996 by Statoil, the Norwegian state-owned oil company, at one of its natural gas fields.

Schlumberger expressed interest but wanted expertise from the petroleum production sector, so suggested approaching ExxonMobil as a potential partner. ExxonMobil agreed to join, but only if the project had a much larger reach. It then played the dominant role in bringing GE and E.ON on board.

By this point, the roles had reversed: ExxonMobil was now proposing a project to Stanford. ExxonMobil is a controversial company, vigorously opposed by environmentalists and many scientists working in the climate-change field. How Stanford handled its side of the ensuing negotiations is interesting, as one might have expected a project sponsored by ExxonMobil to have raised intense scrutiny, interest and debate. But the negotiations were carried through at the level of the dean’s offices, with only a few other Stanford faculty members being

consulted. Thus the deal slid through the approval process, virtually unnoticed, with minimal faculty and no student involvement.

Many universities have an independent research board to provide an oversight of such arrangements with industry, but Stanford has no such body. It is an optional courtesy for members of the faculty senate to be informed of projects of this scope and nature, but on this occasion it did not happen. When the project was announced to the Stanford community, it was a done deal. The sponsors’ contributions to the G-CEP were set to total \$100 million from ExxonMobil, \$50 million each from GE and E.ON, and \$25 million from Schlumberger, over the next ten years. Franklin Orr was to be director of the programme.

Political agenda

To put the motivations of the G-CEP sponsors in context, one must understand the environment in which they have operated and their responses to it. During the Clinton–Gore years, the US environmental movement was on the rise, and the US administration accepted the view that global warming was a clear and present danger. It proposed that consumption of fossil fuels should be reined in (the Kyoto protocol on climate change); that broad steps should be taken to encourage the use of renewable energy sources in place of fossil fuels; and that energy conservation should be strongly

Stanford University has shown a marked blindness to the political agenda of its sponsors.

P. SAKUMA/AP

encouraged. Naturally, the the automobile industry and major industries associated with the provision and consumption of fossil fuels felt threatened and denigrated by these proposals.

ExxonMobil played a leading role in orchestrating and bankrolling the opposition. It is the world's largest oil company, with annual revenues exceeding \$210 billion, and wields enormous political power globally. Its chief executive, Lee Raymond, is a dedicated believer in the free market. In an interview with David Buchan and Sheila McNulty in the *Financial Times* published on 12 March last year, Raymond is quoted as stating: "In Europe you like to tell people what kind of cars they ought to use. Most Americans like to make that decision themselves and that's why they left [Europe]."

The campaign led by ExxonMobil was waged on several fronts, questioning whether global warming was a danger and arguing that any government restrictions imposed on the supply or use of fossil-fuel energy would severely damage US and world economies. The campaign maintained that technological solutions to the greenhouse-gas problem would appear and eliminate the need to regulate fossil-fuel consumption. To accomplish these aims it selectively bankrolled scientists who agreed with its contentions, funded various groups supporting its objectives, lobbied tirelessly and expensively in Washington, and campaigned extensively using full-page 'edu-adverts' in major newspapers. In the process, it left the environmental movement apoplectic.

Since the election of the Bush–Cheney administration, the tables have turned. Now the environmental campaigners are outsiders and the views and objectives of the fossil-fuel industries, as exemplified by ExxonMobil, are mainstream US government policy. For instance, on the same day

that the G-CEP was announced, James Connaughton, chairman of the White House Council on Environmental Quality, addressed a conference of the American Petroleum Institute, an industry-wide association chaired by Lee Raymond. He reasserted the Bush administration's rejection of the Kyoto protocol, arguing that technological innovation, not mandatory curbs, was the key to solving climate-change, telling his audience "climate change is a technology issue".

ExxonMobil reaffirmed its full commitment to fossil fuel at its annual shareholders meeting on 29 May last year, by simultaneously announcing its plans to spend \$100 billion over the next decade on oil exploration and defeating a stockholder's motion for the company to start playing a part in the renewable energy field. Surprisingly, the renewable-energy motion got 20% of the vote, up from the 9% the previous year. Regarding his company's opposition to entering the field of renewables, Raymond is cited in an interview with Thaddeus Herrick of the *Wall Street Journal* on 30 May last year: "If our main competition wants to invest in solar, God bless them."

G-CEP restrictions

So how does this affect the G-CEP? Its director, Franklin Orr, is independently appointed by Stanford with full control of the day-to-day management of the programme. An advisory board, comprised of members independent of either the sponsors or Stanford, will be appointed. It is hoped that this board will be appointed solely on the basis of scientific qualifications.

Less satisfactorily, the sponsors have the right to withdraw from the G-CEP with three years' notice. Originally, the project was to run for a minimum of ten years. A shorter guaranteed term is not compatible with the G-CEP's aim to investigate long-term projects, for which the sums of money involved will require new faculty members, support staff and graduate students. Termination of G-CEP funding would result in painful readjustments or layoffs for staff. Thus the sponsors have powerful leverage over Stanford.

The sponsors appoint the members of the management committee. Research projects will be proposed by Stanford to the management committee, which will accept or reject them. Once accepted, projects will then be run by Stanford. Already, the G-CEP is oriented towards the sponsors' agenda. Despite the sponsors' obvious aversion for conservation and renewable energy sources, Stanford has acquiesced, giving the sponsors real power to set an agenda for the programme.

All patents resulting from Stanford's research will ultimately be vested in Stanford. But the sponsors will be granted first-use rights for five years, free of royalties. These restrictive provisions give the sponsors a clear edge over their competitors and



Lee Raymond of ExxonMobil, a company committed to expanding fossil fuel use.

are out of line with traditional academic practices.

Undeniably, there are strings attached to the G-CEP. The sponsors stand to reap enormous political and public relations advantages through their sponsorship, at a cost that is minuscule compared with their scale of operations. The yearly price tag to ExxonMobil, the major contributor, for example, is merely a third of what it pays its chief executive.

Stanford's administration seems to have been a compliant negotiator in their eagerness to land the project, failing to recognize the strengths of the university's position. Instead, it is relying on the goodwill of the sponsors and the toughness of the project managers to ensure the integrity of Stanford's involvement. The Stanford administration seems to have forgotten the fundamental rule of funding: "If you don't want to be jerked around, don't let your sponsors tie strings to you."

Funding of universities by industry is rapidly increasing in the United States. The concessions that Stanford has made were not needed, and could potentially compromise its academic independence. Other universities entering into such arrangements in the future should be alert to the pitfalls inherent in the Stanford model, and should avoid being tempted by large donations into acquiescing to inappropriate restrictions. ■ David Ritson is at the Varian Physics Department, Stanford University, Stanford, California 94305-4060, USA

Acknowledgements

I thank Peter Ritson for advice and help.

The G-CEP

♦ gcep.stanford.edu

Schlumberger and the G-CEP

♦ www.slb.com/Hub/index.cfm?id=id1357148

ExxonMobil and the G-CEP

♦ www2.exxonmobil.com/corporate/Newsroom/NewsReleases/xom_nr_201102.asp



Franklin Orr: keen on collaboration.

Knowledge is power

Self-interest shaped the global agreement on intellectual-property rights.

Information Feudalism: Who Owns the Knowledge Economy?

by Peter Drahos & John Braithwaite
Earthscan: 2002. 254 pp. £35 (hbk); £12, £10.80 online (pbk)

Sandy Thomas

Access for the poor and sick to patented medicines for infectious diseases. 'Pirating' of software and music CDs. Traditional knowledge in developing countries. These issues command the attention of both the media and politicians in the world's leading economic fora. They are all regulated by TRIPS (the Agreement on Trade-Related Aspects of Intellectual Property Rights), which was probably the most important agreement for the protection of inventions in the twentieth century. Its legacy is all around us.

One of the 28 agreements that made up the Uruguay round of the General Agreement on Trade and Tariffs (GATT), TRIPS was signed by more than 100 countries in 1994. In essence, it extends the standards for the protection of intellectual property from developed countries to the developing world. The adoption of patent systems and copyright protection were agreed by emerging economies such as those of India, China and Brazil, which are masters of electronic and pharmaceutical copying, as well as by some of the world's poorest countries, which have filed very few, if any, patents.

Nearly ten years on — as the developing world faces the ravages of HIV, malaria and tuberculosis, reducing life expectancy from 65 to 39 years in Zimbabwe, and as increases in crop yields level off — tough questions are being asked about the economic and moral obligations of the richer countries to share their intellectual and financial capital. The advance of the 'knowledge economy' in industrialized and emerging nations means that having a competitive advantage increasingly depends on the management and understanding of information.

Arguments over the freedom of researchers, entrepreneurs and industrialists to use such information are not new. They have been fiercely debated in the UK Parliament, the US Congress and other legislatures for over two centuries. Perspectives on the use of scientific knowledge have depended in part on what we seek to do with it. We may argue that the more widely scientific knowledge is used, the greater is its value, because each user gains something at little or no cost. As a result, society will benefit from this public good. But many products that incorporate scientific knowledge, including medicines, can easily be copied at a fraction



Money for nothing? Rich countries' power over knowledge assets is similar to the charging of rent by powerful land-owners, shown here in Pieter Bruegel the Younger's *The Rent Collectors*.

of the cost, thanks to savings in R&D and marketing. The need to provide a system of incentives to investors in the companies that originated a product has led the development of complex systems of legal protection.

Today's global agreements have been shaped by the way in which past governments and industries have sought to control the use of knowledge in the context of innovation to protect national and corporate interests. In *Information Feudalism*, Peter Drahos and John Braithwaite paint a deeply unflattering picture of the behaviour of major corporations over the past century. They describe how US corporations in particular, no longer able to secure monopolies by operating lucrative cartels in commodities because of the introduction of antitrust laws, have turned to the patent system to control and segment markets.

Drahos and Braithwaite argue that developments over the past 30 years are driving us dangerously towards a system that they refer to as information feudalism. They draw parallels between medieval feudalism, with its inequality of property ownership, and the exercise of monopolistic power over knowledge assets.

Central to this analysis is the account of the negotiation of TRIPS, whereby the campaign for globalized intellectual-property standards was shifted to the international trade agenda. Developing countries were persuaded to sign up to TRIPS in exchange for the liberalization of world trade markets. The subsequent failure of these markets to materialize (witness US steel tariffs and farm subsidies in the United States and Europe)

also goes some way to explaining the growing disenchantment with TRIPS.

The authors undoubtedly have a story to tell. Many of the historical facts make interesting reading. The treatment meted out to early 'pirates' — an astonishing 40% of the inmates in the Bastille during the 1750s had offences connected to the book trade — and the complex history of patents and copyright and their effects on free trade form the backdrop for the globalization of corporations and the rise of the information age in the twentieth century. This book is not a technical account of the world's intellectual-property systems; rather, it is about the exercise of power and politics that led to the adoption of TRIPS.

Drahos and Braithwaite are intensely critical, indeed scathing, about the alleged motives of multinationals, particularly in the pharmaceutical and IT sectors, in getting the issue of global standards for intellectual-property rights onto the global trade agenda. The chapter that describes the United States' aggressive behaviour in investigating, by means of its 'Special 301' list, and applying bilateral pressure to those countries that did not conform to its standards of intellectual property, will raise the eyebrows of many a reader.

Much of the history of intellectual-property rights is, to say the least, ironic. Britain, much of continental Europe and the United States have over the past decade supported the global extension of their intellectual-property standards in an effort to stamp out the widespread copying of technology in the developing world. Yet these same countries

were themselves guilty of using these strategies in the past to further their own economic interests. Until 1891, copyright protection in the United States was restricted to US citizens and, between 1790 and 1836, patents were similarly restricted. The Netherlands and Switzerland also avoided adopting a patent system when industrialists wanted to make use of foreign inventions.

This is not a book for the faint-hearted. General readers may wish to understand why patents and copyright have become prominently linked to such issues as access to medicines in poor countries and the consequences of pirating software and music. But they may become weary as the book traces, often in great detail, how representatives from many industrial, government and other organizations allegedly conspired to manipulate patent and copyright systems to become accomplished operators in the 'knowledge game', to the detriment of the public interest and the developing world.

Those who stay the course will gain a clear insight into why so many non-governmental organizations are furiously lobbying for the removal of TRIPS and for reform of the patent and copyright systems. But they will not discover why many other constituencies, including industry, universities and governments, are broadly in favour of protecting intellectual property. Crucially, evidence of why the patent system has, on balance, almost certainly benefited consumers hardly gets a mention. As a consequence, many of the arguments advanced in the book are seriously flawed.

Neither do the authors get to grips with some of the most important questions in development policy that loom large today: notably, how should the moral and economic responsibility of addressing the burden of disease and food insecurity that affects developing countries be apportioned, and by whom?

Sandy Thomas is at the Nuffield Council for Bioethics, 28 Bedford Square, London WC1 3JS, UK. The authors' views are her own

Science laws and science lore

Cassell's Laws of Nature: An A-Z of Laws and Principles Governing the Workings of Our Universe/The Nature of Science: An A-Z Guide to the Laws and Principles Governing Our Universe

by James Trefil

Cassell Reference/Houghton Mifflin: 2002/2003. 433 pp. £25/\$35

Walter Gratzer

After he pulled off, with his *Brief History of Time*, the biggest publishing coup since the Bible, Stephen Hawking revealed that he owed it all to a sage injunction from the publisher: no equations, for every equation would halve the sales. $E=mc^2$ alone was permitted. James Trefil's text is fairly liberally adorned with equations, but they are more for show than for instruction, because his primary target is the lay audience.

Trefil has assembled a multitude of laws, rules, principles and maxims of science. I had not previously heard of Cope's law — though I am willing to believe that where palaeontologists gather they speak of little else — or Allen's rule, or many of the others that feature in Trefil's 235 short essays. His book is, in truth, an encyclopaedia, encompassing the whole panorama of science, with potted biographies of celebrated protagonists, from Archimedes and William of Occam to the heroes of today.

Trefil takes the view that anything too rarified or obscure for the standard text-

books does not warrant inclusion. There is, as he concedes, room for argument around the edges: why, for instance, Graham's law of the diffusion of gases, but not the principles of colligative solution properties, such as Raoult's law? The phase rule is out, as is the Stark effect, but Zeeman is in.

Biology tested pretty well for comprehensiveness. Trefil even finds room for some defunct generalizations, such as the rule that ontogeny recapitulates phylogeny (which he quite properly rejects), yet homeotic genes — among the most important discoveries in twentieth-century biology — are an odd omission.

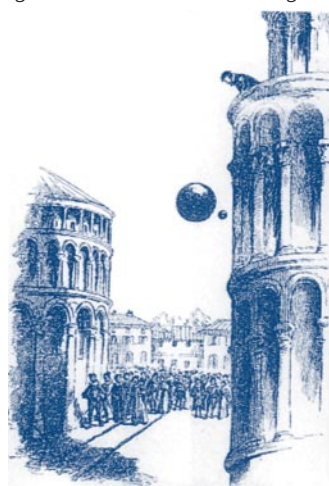
Still, there is plenty here to nourish all but the most omniscient reader. Trefil writes in an easy, colloquial style, even if some of his analogies are a trifle laboured. His

great virtue is that he does not funk the most daunting subjects in the physical sciences; indeed he generally prefaces his expositions with soothing reassurances that the matter is not as difficult as it may seem. For the most part, he adheres to the excellent principle that a little inaccuracy often saves a world of explanation.

In his dissertation on Heisenberg's uncertainty principle, he suggests that if your mind rebels at such an affront to common sense, you should ask yourself: "Why

not? How do I know what things are supposed to be like inside the atom? Have I ever been inside an atom?" This strikes me as a mite facile, for the overthrow of determinism did, after all, trouble many of the best minds of the twentieth century. It was Max von Laue and Otto Stern who swore an oath (broken, to be sure) that "if this nonsense of Bohr's turns out in the end to be correct, we will leave physics".

Einstein never renounced his belief that "God does not play dice" (countered more recently by Stephen Hawking, who declared that not only does God play dice, but he sometimes throws them where they can't be found). In cleaving nevertheless to pictorial representations, Trefil chooses to discuss not only atomic physics, but also chemical bonding in terms of the archaic Bohr-Rutherford planetary atom, with



New in paperback

The Evolution Explosion: How Humans Cause Rapid Evolutionary Change

by Stephen R. Palumbi

W. W. Norton: 2002. 288 pp. \$14.95, £11.95

"Those not besotted by the anchovy daiquiris of creationism will be convinced by Palumbi's book that evolution is alive and well." Jerry A. Coyne, *Nature* **412**, 586–587 (2001).

Einstein's Unfinished Symphony: Listening to the Sounds of Space-Time

by Marcia Bartusiak

Berkley: 2003. 272 pp. \$14

The Biomechanics of Insect Flight: Form, Function, Evolution

by Robert Dudley

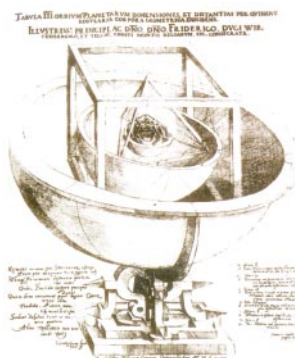
Princeton University Press: 2002. 536 pp. \$35, £24.95

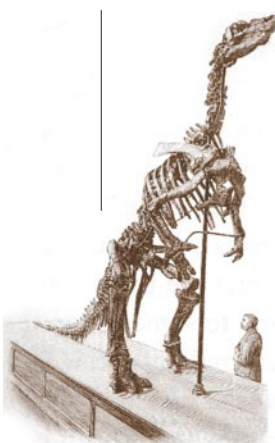
"Robert Dudley has written a remarkably comprehensive account of our knowledge of insect flight." R. McNeill Alexander, *Nature* **405**, 17–18 (2000).

Nearest Star: The Surprising Science of Our Sun

by Leon Golub & Jay M. Pasachoff

Harvard University Press: 2003. 267 pp. \$16.95, £11.50, £16.95





only the sketchiest allusion to atomic orbitals. The problem remains that, as one physicist has put it, we have no words for the quantum world.

Trefil is a physics professor who served time as a high-energy physicist, and his entries in the physical sciences are accordingly the most authoritative. There are some curious errors nevertheless: what, for instance, is the Rydberg constant doing in the Clausius–Clapeyron equation? Was this introduced (in place of the gas constant) by the errant hand of a copy-editor? The shear elastic modulus is wrongly defined, the description of Oersted's famous experiment misses the point, and the section on spectroscopy recognizes only the electronic sort, with not a mention of vibrational, nuclear magnetic or electron-spin-resonance spectra.

The biological topics are for the most part clearly laid out, but again they are not altogether devoid of errors. Most are trivial, but blue-eyed children are not born to brown-eyed parents, nor does a B-cell display many kinds of antibody on its surface; the experiments of Miller and Urey and their successors on the origins of life have never given rise to anything that can decently be called a protein or a lipid; bacteriophages do not all contain DNA, nor, when they do, is it necessarily stitched into the host genome; and there are 20 protein amino acids.

The material is abundantly cross-referenced, yet there is a good deal of repetition. A few entries, says Trefil, are “just plain fun”; so there is the ‘beauty criterion’, Occam's razor and Murphy's law (although without the various well-known corollaries — when you want to do something, there is always something else you have to do first, and so on). Fermat's last theorem surfaces, presumably because it has of late entered into public consciousness (as reflected by the graffiti in the New York subway: “I have a beautiful proof of this theorem but can't give it here because my train is coming”).

This is a handsomely produced and stylishly illustrated volume, and should make a helpful and attractive basic reference source. I am now better informed on several topics of which I knew little before for having read it. So more power to Trefil for a brave and largely successful effort. ■

Walter Gratzer is at the Randall Centre for Molecular Mechanisms of Cell Function, King's College London, New Hunt's House, Guy's Campus, St Thomas Street, London SE1 1UL, UK.

Mergers and acquisitions

Darwin's Blind Spot: Evolution Beyond Natural Selection

by Frank Ryan

Houghton Mifflin/Texere: 2002/2003. 320 pp. \$25/£18.99

Steven A. Frank

Photosynthesis arose in several independent bacterial lineages. Traditional evolutionary analysis would suggest that photosynthetic biochemistry originated independently in each of the separate lines of descent. However, recent genomic comparisons between five groups of photosynthetic bacteria show that there has been widespread horizontal gene transfer. Several bacterial groups acquired key components of photosynthesis by getting genes from other lineages.

The plant *Dichanthelium lanuginosum* grows in geothermal soils that get very warm. Like nearly all land plants, *Dichanthelium* has symbiotic fungi. These fungi enhance the growth of *Dichanthelium* at temperatures below 40 °C but are essential above 40 °C; plants without symbiotic fungi die at higher temperatures. So colonization of geothermal soils by *Dichanthelium* requires symbiosis.

Retrotransposons make up about 38% of mouse and human genomes. Those transposons presumably invaded ancestral genomes by retroviral infection. It is not yet clear whether retrotransposon DNA contributes significantly to host characters. However, the origin of some important host characters will probably be traced to this vast genomic component descended from viral genes.

According to *Darwin's Blind Spot*, if you are shocked by these observations, you are a

true darwinian; if not, you are a radical symbiologist. Frank Ryan's darwinians believe that evolutionary change can arise only by descent with modification within clearly defined lineages. The radical symbiologists believe that major evolutionary innovation often arises by the joining of genetic information from different lineages.

The three observations on symbiosis quoted above come not from Ryan's book, but from a few recent issues of *Science* and *Nature* that I happened to read last night. Must we conclude that these journals have abandoned traditional darwinism and quietly gone radical?

Ryan has built an exciting story of heroic outsiders and fierce conflict over the nature of evolutionary innovation. There were indeed mighty battles over the origin of mitochondria and the eukaryotic cell, but they ended decades ago. Now everyone accepts that mitochondria descended from an independent bacterial lineage, and that the eukaryotic cell has symbiotic origins.

I agree with Ryan that mainstream evolutionary biologists still often fail to consider symbiosis as a plausible hypothesis to explain puzzling characters. But this failure does not arise from deep convictions about the nature of evolution. Rather, the mainstream comes round to a new way of thinking only after the evidence has piled up. Genomics will help greatly here because it allows us to trace the evolutionary history of lineages and untangle the complex web of descent.

Studies of symbiosis will surely lead to great progress in understanding ecological interactions and evolutionary history. On this most important point, Ryan is right. So why is it necessary to have heroes and villains, and to portray mainstream science as hopelessly conservative and plainly wrong?

Sales. Ryan is a successful author who



On the move: the broken, variable colour of dahlias is caused by transposons, or ‘jumping genes’.

J. BURGESS/SPL

knows what he is doing. The story moves along. Anecdotes fascinate, personalities excite. Reviewers for www.amazon.com loved his earlier book, *Virus X* (Little, Brown, 1997). The style owes much to recent popular literature in biology. Come into my living room for a few hours and let me regale you with stories, he says. It'll be fun. And in the end, you'll know why most scientists understand little of the wonders of nature — they are blinded by their conservatism and prejudice. So, just by listening to me, you'll end up smarter than the professionals. It feels nice.

Ryan sometimes takes too broad a view of symbiosis. For example, he emphasizes that humans depend on food sources for certain

essential vitamins and amino acids that we could, in principle, make for ourselves. "Viewed from a Darwinian standpoint, this makes little sense. Viewed from a symbiotic standpoint... we humans have evolved in a diffuse exosymbiotic relationship with the living providers of these 'missing genes'... With the exception of autotrophic bacteria, every life form on Earth has similar exosymbiotic needs." So, if you are a Darwinian, eating puzzles you.

Despite Ryan's caricature of Darwinism, he does understand well enough the basic principles of natural selection. However, his emphasis often creates the impression that symbiosis is the only way to achieve major evolutionary change.

For example, he approves of panspermia, which holds that life came to Earth on a meteorite. Maybe it did, I don't have an opinion. But if Earth acquired life from another source, where did life originate?

This question reminds me of a story about Bertrand Russell. He gave a lecture in which he mentioned some properties of the Solar System. Afterwards, a woman told him that the Sun was held up on a turtle's back. Russell asked what held up the turtle. "You think you are so clever," she replied. "It's turtles all the way down."

Steven A. Frank is in the Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697-2525, USA.

Science in culture

Visual zoology

Historic zoological wall charts found in Pavia.

Alison Abbott

Rudolf Leuckart (1822–98) was a respected zoologist and parasitologist. He lived in the heyday of the 'new zoology', which was influenced by both Darwinism and the new discipline of physiology, and was increasingly scientific in its approach.

At a time when the life cycles of organisms were just starting to be understood, Leuckart was at the forefront. In one notorious experiment, for example, he found a volunteer to swallow four cysticerci, the larvae of the human tapeworm, to prove that these 'bladder worms' would develop into tapeworms using the human as host. One month later, the unfortunate volunteer was duly purged of two 2-metre tapeworms, and the prevailing concept that such

parasites were generated spontaneously was finally laid to rest.

But Leuckart, who became rector of the University of Leipzig in 1877, left just as powerful a legacy in the form of his didactic *Wandtafeln*, or wall charts. The use of large scientific wall charts as teaching aids became popular in universities in the second half of the nineteenth century. The fashion followed the introduction of *Wandtafeln* in German schools as part of broad educational reforms inspired by Johann Heinrich Pestalozzi, who believed that children should see and experience, rather than learn by rote. Zoological wall charts in particular were valued at universities around the world.

Leuckart's beautiful series was neither the first nor the last to be produced, but is arguably the most significant, particularly given its extensiveness. There were 101 wall charts in his series on invertebrates, although the second series, on vertebrates, co-edited by his younger colleague Carl Chun, peters out after just 12.

The artwork was mostly produced by a series of Leuckart's assistants, who inevitably needed artistic as well as scientific skills. Leuckart himself illustrated only one of the charts with his own hand, that of his tapeworm (shown on the left). Each chart presents a single species, depicting its anatomy in different stages of growth or life cycle, and often alluding to function.

Their styles vary somewhat. But the 22 artist-zoologists all owe a debt to their contemporary Ernst Haeckel, a vociferous promoter of Darwin's new theory of evolution, whose superbly composed and beautifully drawn illustrations revolutionized the presentation of zoology. Haeckel used some of his own charts in his public lectures, happily compensating for his reportedly poor verbal skills.

Leuckart's wall charts are not labelled with any explanatory text. Detailed explanations are instead provided in pamphlets intended for teachers, in three languages: German, English and French.

Production of the charts was technologically challenging and labour-intensive. Their size — 1 metre by 1.4 metres — required that they be

printed in four sections, before being joined together and touched up. A complete set cost more than twice as much as the most expensive microscope of the time, yet they sold widely.

Unfortunately, very few of the charts have survived. So the discovery in 1997 of an almost complete collection, with full explanatory texts, at the University of Pavia in northern Italy is exciting. The charts were found when several of the university's science departments moved out of the eighteenth-century Palazzo Botta, where they had been situated for more than two centuries, into modern laboratories. Cleaning out long-forgotten attic and cellar stores, zoologist Carlo Redi stumbled across a large, 18-drawer chest in which the charts had been collecting dust for decades. The invertebrate series is complete, and only three charts in the vertebrate series are missing; these have since been located in Hamburg.

Only one other nearly complete series with explanatory texts is known to exist. It is held at the Marine Biological Laboratory in Woods Hole, Massachusetts. Again the invertebrate series is complete, but four charts are missing from the vertebrate series. Some 40 individual charts have been located in Hamburg, Leipzig, Rome and Padua.

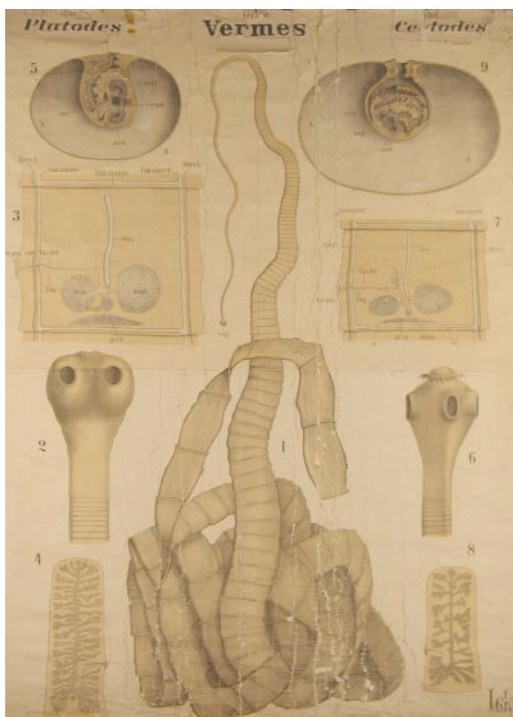
The Woods Hole collection is available digitally on the laboratory's homepage. The Pavia collection has also been digitized and has just been published as a sumptuous book, *Visual Zoology*, with an accompanying CD. Redi hopes to make the collection available on the University of Pavia's homepage soon.

Alison Abbott is Nature's senior European correspondent.

The Pavia collection can be seen in *Visual Zoology: The Pavia Collection of Leuckart's Zoological Wall Charts, 1877*, edited by Carlo Alberto Redi, Silvia Garagna, Maurizio Zuccotti, Ernesto Capanna & Helmut Zacharias (Ibis, 2002).

Woods Hole collection

► <http://www.mbl.edu/leuckart>



A questioning mind

Summarize yourself in the form of a title of a paper in Nature.

Passionate curiosity leads to endless frontiers.

What was your first experiment as a child?

My first observation was on silkworms. I was amazed to watch the dramatic changes from egg to larva — a worm that eats only leaves but spins two kilometres of silk — to the moth that emerges from the cocoon to lay more eggs.

Who has been the most important mentor in your career?

Alexander Rich at the Massachusetts Institute of Technology (MIT), who has advanced our understanding of nucleic acids and protein synthesis. His office door has always been open, and I now follow his example.

What makes a good scientific mentor?

The capacity to inspire and give freedom to students and postdocs, to let them explore and make discoveries themselves.

Whose graduate student would you most like to have been (historical impossibility notwithstanding)?

Francis Crick. He is always full of great ideas and restless in intellectual pursuit. He moves from one field to another effortlessly, is very stimulating and is not afraid to ask the big questions that most others do not.

What single scientific paper changed your career path?

Alexander Rich and colleagues reported the discovery of left-handed DNA (A.H.-J. Wang *et al. Nature* **282**, 680–686; 1979) when I was an undergraduate in China. The previous year I had asked my biochemistry professor why all biological helices seemed to be right-handed, and whether there might be left-handed ones. My professor did not know, but that single paper led me to pursue a career with Rich. After many years of learning English and molecular genetics at the University of California, Santa Barbara, I joined his lab at MIT as an American Cancer Society postdoctoral fellow, studying the biological function of left-handed DNA, and this led to my own serendipitous discovery.

What book has been most influential in your scientific career?

One, Two, Three ... Infinity by George Gamow, the great physicist who also proposed the Big Bang theory and the three-letter genetic code. Gamow was a visionary, proposing grand ideas most people thought were crazy. Although not all of his ideas

turned out to be correct, his most important ones have stood the test of time.

What literary character would you employ as a postdoc?

James Bond. He is multi-talented, witty, intelligent, smart, resourceful and fearless. He could move into any research field without hesitation, and would get interesting and exciting results almost single-handedly!

What's your favourite conference destination, and why?

Crete, location of the technologically advanced but mysterious Minoan civilization, and more recently of the biennial workshop on the self-assembly of peptides and proteins in biology, medicine and engineering, an interdisciplinary gathering of scientists and engineers. Francis Crick put it best: in nature, hybrid species are usually sterile, but in science the reverse is often true. Hybrid subjects are often astonishingly fertile, whereas if a scientific discipline remains too pure it usually wilts.

You have the audience in your hands, but some smart-alec asks you the killer question you have no idea how to answer. What's your favourite response?

"We have thought about it, but put it aside temporarily because of a shortage of manpower."

What book currently resides on your bedside table?

Mendeleyev's Dream: The Quest for the Elements by Paul Strathern, a book about the history of science, especially chemistry and medicine.

What music heads the playlist in your car or laboratory?

Mozart, of course! His music is so pure and elegant, even though he composed it amid many distractions and often at the last minute. He was a musical magician!

Assuming the dead can be raised and/or time travel exists, who from the world outside science would you most like to have dinner with?

Mozart, Beethoven, Chopin and Tchaikovsky. I would love to know what inspired them to compose their timeless music.

Where and when would you most like to have lived or worked?

The Cavendish Laboratory in Cambridge in the early 1950s. People then had more time to think deeply about timeless questions, many of which remain unanswered today. Now people work too much and think too little,



Shuguang Zhang

Shuguang (Chinese for 'daybreak') is at the Center for Biomedical Engineering at the Massachusetts Institute of Technology (MIT). He is curious about many things: in 1994, a serendipitous discovery of self-assembling peptides was selected to be one of the 15 achievements of the past 25 years at MIT.

always in a hurry but less concerned about originality, the wellspring of new fields.

You are on a plane behind two students obviously going to the same conference, who start to talk about your work. What do you do?

I would introduce myself and ask them how they can push it further.

What's the best piece of advice you've ever received?

Received from Francis Crick: always ask questions, the bigger the better. If you ask big questions, you get big answers.

What single discovery, invention or innovation would most improve your life?

Three-dimensional cell culture, and the ability to build a tissue or organ from stem cells and a peptide/protein scaffold.

Name one extravagance you can now get away with because of your eminence.

I can invite the most well-known people out for dinner.

What music would you have played at your funeral?

Five Mozart piano concertos: No. 6 (K. 238); No. 9 ('Jeunehomme', K. 271); No. 23 (K. 488); No. 26 ('Coronation', K. 537) and No. 27 (K. 595).

What's just around the corner?

Construction of nanodevices from the bottom up, using biological molecules as nanoscaffolds. ■

The cost of living

Peter Rich

The 1960s was an important time for mitochondrial research, which saw many significant and enduring advances. During this period, Peter Mitchell developed the chemiosmotic proposal — the coupling of biological electron transfer to ATP synthesis. Although it ultimately won Mitchell a Nobel Prize, his work was highly controversial, and it was more than a decade before his ideas were accepted widely, leading some researchers to claim prior ownership of key elements of the ideas that still persist today.

The fact that mitochondria are central to the efficient provision of energy for eukaryotic cells is undisputed. It is generally accepted that the early and energetically inefficient eukaryotic cell was invaded more than a billion years ago by a bacterium containing a much more efficient system for utilizing available energy sources — the oxygen-consuming respiratory chain. The majority of the bacterial genetic information was subsequently transferred to the nucleus, transforming the bacterial symbionts into slave mitochondrial organelles.

We rely on these mitochondria to synthesize ATP. The process begins with the passage of electrons derived from food along a series of mitochondrial respiratory-chain carriers, until they are consumed in converting the oxygen we breathe into water. Mitchell showed that their transfer is linked to the movement of protons across the ion-impermeable inner

mitochondrial membrane in which the respiratory chains are embedded. The membrane acts as a capacitor, storing energy as a pH and charge difference across the membrane. This gradient of electrochemical potential energy is used to drive protons through the ATP synthases that are embedded in the same membrane, and which couple the exergonic proton flow to the synthesis of ATP from ADP and phosphate. The energy thus stored can be released by ATP hydrolysis, a reaction that is used by the myriad energy-requiring enzymes that maintain cellular function.

An average human at rest has a power requirement of roughly 100 kilocalories (420 kilojoules) per hour, which is equivalent to a power requirement of 116 watts — slightly more than that of a standard household lightbulb. But, from a biochemical point of view, this requirement places a staggering power demand on our mitochondria. Mitchell's work showed that the electrochemical gradient of protons across the inner mitochondrial membrane that drives ATP synthesis is roughly 200 mV, and most of this is the electric field component.

If it is assumed that 90% of human power is provided by the protons that are transferred through the ATP synthase, then the transmembrane proton flux would have to represent a current of 522 amps, or roughly 3×10^{21} protons per second. Recent estimates, based on the structure of yeast ATP synthase, contend that three ATP molecules are formed for every ten protons that are transferred. Assuming a conversion efficiency that is close to unity, ATP is reformed at a rate of around 9×10^{20} molecules per second, equivalent to a turnover rate of ATP of 65 kg per day and with much higher rates than this during periods of activity. This output is itself powered by the oxygen-consuming respiratory chain.

A typical adult male consumes around 380 litres of oxygen each day, and top athletes can sustain rates that are ten times greater for limited periods. Most (90%) of this oxygen is reduced to water by the terminal respiratory-chain enzyme, cytochrome oxidase. The inner mitochondrial membrane contains around 0.4 nanomoles of this enzyme per milligram of protein. It can work at a rate in excess of 300 electrons every second, but probably operates at an average rate of no more than 50 per second. Hence, an average human will need 2×10^{19} molecules of cytochrome oxidase to support oxygen consumption. With the inner mitochondrial membrane having a lipid/protein weight ratio of 1:1, the cytochrome oxidase would be associated with about 70 ml of lipoprotein membrane. However, the membrane's thickness — only 6 nm — means that the surface area of the inner mitochondrial

Chemiosmotic coupling

We may well be paying a significant price for the luxury of an efficient energy supply.

membrane in an average human would be around 14,000 m².

This constant energy provision is a herculean task so it is not surprising that defects in mitochondrial function should lead to physiological disorders. As human mitochondrial DNA (mtDNA) sequence and clinical databases are amassed and compared, the number of diseases that are suspected to be caused by mitochondrial dysfunction due to mtDNA mutations is increasing. Mutations in mtDNA might even accumulate with age and contribute to a reduced efficiency of energy provision. Some aspects of mitochondrial function seem to be potentially deleterious to cell health, particularly the 'leakage' of electrons that must inevitably occur, resulting in the formation of small amounts of damaging and mutagenic free radicals. Mitochondria may also play a key role in some cells in an apoptotic, or 'cell death', pathway by releasing factors from the intermembrane space into the cytoplasm, where they interact with the caspase cascade system. So it seems that our energy-producing organelles have not yet been entirely tamed to the needs of the eukaryotic cells that control them, and that we may well be paying a significant price for the luxury of the efficient energy supply that we need to sustain our ravenous needs.

Recognition of such mitochondrial activities, together with the tremendous advances in determining the structural details of the mitochondrial molecular machinery, in many instances down to the atomic level, is heralding a new era of interest in mitochondria. Hopefully, the new wave of insights into this enigmatic organelle that are now emerging will not evoke the same controversy and prolonged difficulties in acceptance as Mitchell's work. ■

Peter Rich is in the Department of Biology, University College London, Gower Street, London WC1E 6BT, UK.

FURTHER READING

Gilbert, G. N. & Mulkay, M. *Opening Pandora's Box* (Cambridge Univ. Press, 1984).
Prebble, J. & Weber, B. *Wandering in the Gardens of the Mind: Peter Mitchell and the Making of Glynn* (Oxford Univ. Press, New York, 2003).
Catalogue of the Papers and Correspondence of Peter Dennis Mitchell FRS NCUS (Cambridge Univ. Libr., Cambridge, UK, 1997).
Kiberstis, P. A. *Science* **283**, 1475 (1999).

THE ART ARCHIVE/ARCHAEOLOGICAL MUSEUM NAPLES/DAGLI ORTI



Shining example: even at rest, a human body requires as much power as a 100-watt lightbulb.

Parasites lost

Keith Clay

Why do some plants and animals become pests when they are introduced to new areas? Part of the answer seems to be that they have left most of their parasites behind, gaining vigour as a consequence.

Invasive species can be a real bother. These are plants or animals that, when they are accidentally or deliberately moved from one region to another, flourish to the extent of getting out of hand and becoming pests in their naturalized environment. They tend to reduce biodiversity, and can have adverse effects on human well-being^{1,2}. Much effort is devoted to controlling them after they are established, but a better understanding of why species become invasive offers the possibility of taking pre-emptive measures.

In companion papers on pages 625 and 628 of this issue, Mitchell and Power³ and Torchin *et al.*⁴ illuminate the biology of invasiveness. They report the results of surveying parasite loads of invasive plants and animals in their naturalized and native ranges. They find that parasitism is significantly reduced in organisms in the introduced range, so supporting the 'enemy release hypothesis' — the idea that species are more likely to become invasive when they are released from control by their natural enemies.

This hypothesis is one of several (not mutually exclusive) explanations put forward to account for why some species increase uncontrollably when introduced into new areas. In *The Ecology of Invasions by Animals and Plants*⁵, Charles Elton favoured a more general form of this hypothesis, the 'ecological resistance hypothesis', which proposes that competition with native species and attack by native predators and pathogens keep invaders in check^{6,7}. Island systems such as those of Hawaii or Australia may be particularly prone to invasion, for several reasons — their lower species diversity and population density; the absence of particular groups (placental mammals in Australia, for instance, and snakes in Hawaii); and the lack of coevolutionary history with invaders, in which host and parasite have had a long time to coadapt to each other and have reached a stable state⁸. This lack of coadaptation is dramatically illustrated by the lack of resistance of native species to introduced pathogens. More generally, communities with high diversity may be more resistant to invasions⁹.

The ecological resistance hypothesis suggests why certain habitats might be more or less prone to invasion, but does not predict why certain species are more or less likely to invade that habitat. In particular, this hypothesis cannot provide explanations

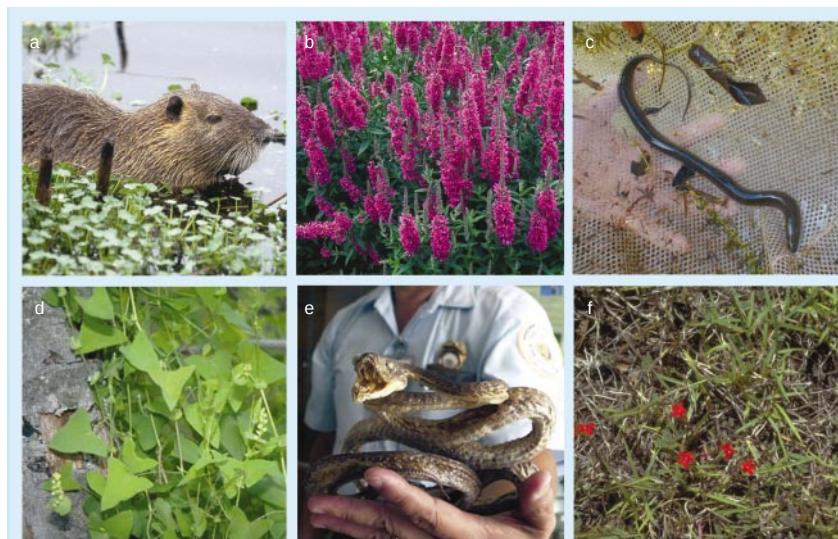


Figure 1 Invasive species in the United States. a, b, Two species that have caused widespread ecological damage to native communities are the nutria, or coypu (*Myocastor coypus*), and purple loosestrife (*Lythrum salicaria*). c, d, Other species such as the Asian rice eel (*Monopterus albus*) and mile-a-minute weed (*Polygonum perfoliatum*) are limited in their distribution but threaten to become severe problems as they expand into areas to which they are well adapted. e, f, Examples of a worrisome group of species that could prove serious pests should they arrive are the brown tree snake (*Boiga irregularis*), which decimated bird life on Guam following its introduction from Indonesia but has not yet successfully invaded Hawaii, and witchweed (*Striga* spp.), which in parts of Africa has a highly adverse effect on cereal yields.

for continental habitats where species diversity is high and all ecological niches seem to be taken, yet where invasive species are numerous. A good example is the eastern United States, where nutria, an aquatic rodent, and purple loosestrife, a wetland plant, are just two of many introduced species that have had or could have adverse effects (Fig. 1). Mitchell and Power³ and Torchin and colleagues⁴ provide a more mechanistic explanation of why these species are problematic.

Both groups compiled published information on parasitism of invasive species in their native and their naturalized ranges. Mitchell and Power³ quantified the number of fungal pathogens and viruses infecting 473 plant species that are native to Europe but have become naturalized and invasive in the United States. They found that these species were infected on average by 77% fewer fungal and viral species in the United States than in Europe, strongly supporting the enemy release hypothesis. Interestingly, in their naturalized range, invasive plants

were more likely to be released from control by fungal pathogens than by viruses.

Mitchell and Power also quantified the characteristics of each plant species from the frequency of reports naming that species as noxious (defined as a weed that poses a high risk to agriculture) or an invader of natural areas. Species that ranked higher on the scales of noxiousness and invasiveness showed greater release from their parasites. The authors found that invasive plant species do accumulate parasites from their naturalized range, but the overall parasite load tended to be less than in their original home. Finally, they also found that invasive plant species with higher rates of pathogen accumulation were less likely to be noxious, an observation that again supports the ecological resistance hypothesis. This provides a ray of hope: if invasive species accumulate parasites in their naturalized range relatively rapidly¹⁰, the problems they cause may be temporary.

Torchin and colleagues⁴ examined 26 invasive animal species ranging from mol-

(A) L. RICHARDSON/CORBIS; (B) E. CRICHTON/CORBIS; (C) MOLLUSCAN PICTURES; (D) M. C. PERRY/USGS; (E) RONENZILBERMAN/AP; (F) J. FRV/AA-Z BOTANICAL

Box 1 Tansy ragwort: a case history

The tansy ragwort (*Senecio jacobaea*, pictured) is native to Eurasia but has been introduced into North and South America and the Australian region. In the western United States it aggressively invaded range lands, where it displaced species of forage plant and poisoned livestock with its pyrrolizidine alkaloids. In 1976, aerial photography showed that the species covered more than 12,000 km² in western Oregon. But by 1988 the area infested had been reduced by 60–90% and more desirable vegetation was recovering, a change that can be attributed to the deliberate introduction of

the cinnabar moth (*Tyria jacobaeae*) from France and the ragwort flea beetle (*Longitarsus jacobaeae*) from Italy. Cinnabar caterpillars feed on developing ragwort inflorescences and leaves, and can totally strip a plant of foliage; flea beetles chew holes in the leaves. The decline of ragwort populations clearly followed the establishment of these natural enemies from the ragwort's native range, and ragwort density has since remained low¹². The same group has experimentally confirmed¹² that the introduction of the natural enemies from its home range



was indeed responsible for controlling the ragwort. K.C.

luscs such as the common periwinkle, *Littorina littorea*, to the black rat, *Rattus rattus*. Like most of the cases examined, these species are native to the Old World but have been introduced to the New World. In addition to quantifying the number of parasite species hosted, the authors also analysed parasite prevalence (the percentage of individuals infected within a population) in the native and naturalized ranges. The parasites documented were exclusively helminths — flatworms or roundworms. There is a rich database on parasitic helminths, reflecting their great diversity and their ecological significance.

Torchin and colleagues' results are similar to those of Mitchell and Power, in that an average of 16 parasite species per host species occurred in the native range compared to seven in the naturalized range. On average, only three of the native parasites accompanied the host to the naturalized range. Hosts were also infected by four parasite species from the new environment, again showing that introduced species accumulate new parasites. But overall parasite prevalence was much lower in the host species' naturalized range.

Taken together, these studies^{3,4} constitute further evidence for the enemy release hypothesis. Invasive species appear to suffer from fewer parasites and pathogens in their naturalized range than in their native range; so, presumably, they are more poorly regulated by parasites in their new environment. This provides conceptual support for one approach to biological control of such species, in which parasitic organisms from plants or animals in native habitats are identified and introduced into the naturalized environment (see Box 1). This strategy can, however, prove a double-edged sword when

the biocontrol agent finds native species to be equally, or more, attractive.

Why are parasites and pathogens often left behind when their hosts migrate? There may be several reasons. The colonization of new habitats is often by just a handful of individuals, increasing the likelihood that all are free of infection. In addition, if parasite dispersal occurs independently of the host, two successful migrations are required. Similarly,

Applied physics

Solar cells to dye for

Michael Grätzel

The idea of producing electricity from sunlight is attractive, but in practice the technology to do so is expensive. A new device, moving away from the traditional silicon design, shows promise.

Photovoltaic conversion of solar energy — from photons to electrons — has so far been dominated by solid-state devices, usually made of silicon and profiting from the expertise of the semiconductor industry. That dominance is now being challenged by new generations of photovoltaic cells. On page 616 of this issue, McFarland and Tang¹ present an intriguing embodiment of a converter that is based on light harvesting by dye molecules on a metal surface. Contrary to expectation, their device shows a strikingly high internal quantum efficiency for electric-current generation.

The silicon used in most of today's solar cells must fulfil several tasks. It must absorb sunlight, converting photons into negative- and positive-charge carriers (electrons and holes, respectively); it must transmit an electric field to separate the electrons and holes;

and it must then conduct these carriers to the current collectors. To achieve all this simultaneously, materials of very high purity are needed, and consequently silicon-based solar cells are too costly to compete with conventional means of producing electric power.

In contrast, the device developed by McFarland and Tang¹ has a multilayer structure that physically separates the processes of light absorption and charge-carrier transport (Fig. 1). Photons are harvested by dye molecules adsorbed on the surface of a thin gold film, which in turn rests on a layer of titanium dioxide (TiO₂). Spontaneous electron flow from the semiconducting TiO₂ layer to the metallic gold layer imparts a slight negative charge to the gold, leaving a slight positive charge on the TiO₂. The resulting local electrostatic field creates a potential

Keith Clay is in the Department of Biology, Indiana University, Bloomington, Indiana 47405, USA.

e-mail: clay@indiana.edu

1. Pimentel, D. (ed.) *Biological Invasions: Economic and Environmental Costs of Alien Plant, Animal and Microbe Species* (CRC Press, Boca Raton, 2002).
2. Westbrooks, R. G. *Invasive Plants: Changing the Landscape of America* (Federal Interagency Committee for the Management of Noxious and Exotic Weeds, Washington DC, 1998).
3. Mitchell, C. E. & Power, A. G. *Nature* 421, 625–627 (2003).
4. Torchin, M. E., Lafferty, K. D., Dobson, A. P., McKenzie, V. J. & Kuris, A. M. *Nature* 421, 628–630 (2003).
5. Elton, C. *The Ecology of Invasions by Animals and Plants* (Univ. Chicago Press, 2000; orig. edn Methuen, 1958).
6. Maron, J. L. & Vila, M. *Oikos* 95, 361–373 (2001).
7. Keane, R. M. & Crawley, M. J. *Trends Ecol. Evol.* 17, 164–170 (2002).
8. Carlquist, S. *Island Biology* (Columbia Univ. Press, New York, 1974).
9. Tilman, D. *Ecology* 74, 2179–2191 (1993).
10. Strong, D. R. & Levin, D. A. *Proc. Natl Acad. Sci. USA* 72, 2116–2119 (1975).
11. Ehrlich, P. R. in *Biological Invasions: A Global Perspective* (eds Drake, J. A. et al.) 315–328 (Wiley, London, 1989).
12. McEvoy, P. & Cook, C. *Ecol. Appl.* 1, 430–442 (1991).

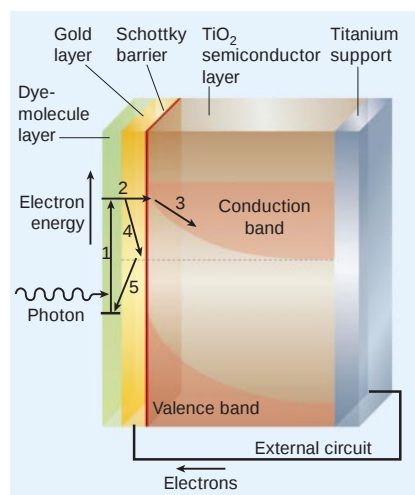


Figure 1 The structure and conversion process of McFarland and Tang's photovoltaic device¹.

Photons hitting the layer of dye molecules cause excitation (1), which results in electrons being injected into the gold layer (2). McFarland and Tang suggest that the electrons move ballistically across the thin gold film and over the Schottky barrier into the conduction band of the TiO₂ layer (3). If instead electrons lose energy in the gold layer (becoming 'thermalized'), they are no longer energetic enough to cross the Schottky barrier (4). An advantage of this device is that the photoexcited dye layer is automatically regenerated by electron donation from the gold film (5).

barrier between the TiO₂ and gold layers, called a Schottky barrier. When light falls on the dye layer, electrons are released from the dye molecules and injected into the conduction band of the metal layer. To generate electric current through the device, these electrons must have enough energy to travel to and over the Schottky barrier, to reach the TiO₂ conduction band. From there, they are transported to the titanium support layer that acts as a current collector, and then to the external circuit. The electron content of the dye layer is replenished by electron donation from the gold film.

The overall efficiency with which this device converts light to electric power is still small (far below 1%), but McFarland and Tang suggest that ultimately it could achieve the same efficiency as an ideal conventional cell. What is remarkable, however, is their finding that about 10% of the photons absorbed by the dye layer actually generate electric current. Dye molecules adsorbed on metals are notoriously ineffective at producing photocurrents². One reason for this is that electrons injected by the dye layer into the metal can be immediately recaptured through reverse charge transfer from filled electronic states in the metal, resulting in no net current flow. Furthermore, the excited dye molecules can be effectively quenched by energy transfer to the conduction electrons of

the metal. The latter process competes with the interfacial electron-injection process.

So why does light-harvesting in this device work so well? McFarland and Tang explain their surprising result in terms of a ballistic pathway for charge transfer across the gold film. The electrons injected by the excited dye molecules are initially 'hot', with excess translational energy. If their motion across the gold film is ballistic (that is, they keep their own momentum), they maintain this energy and hence can cross the Schottky barrier to the TiO₂ layer. But if cooling occurs on the way to the gold–TiO₂ junction, the electrons cannot cross the barrier and do not contribute to the photocurrent. The ballistic model seems plausible in view of the unusually long distance — 20 to 50 nm — that hot electrons can travel in noble metals before their kinetic energy is lost³. There could, however, be another explanation — that electrons are excited in the gold film by energy transfer from the dye layer, and then injected into the TiO₂ conduction band to generate a current.

Human genetics

Lost anchors cost lives

Stanley Nattel

Mutations in ion-transport proteins can destabilize the electrical activity of the heart, causing sudden death. It now seems that mutations in a protein that anchors ion transporters to cell membranes can have the same effect.

Many excitable tissues — including the heart, brain and nervous system — are controlled by the flow of electrical currents across cell membranes. The heart, for instance, beats about 100,000 times a day, with each beat requiring an appropriate and coordinated pattern of muscular contraction that is triggered by finely tuned, rhythmic electrical activity. When the coordinated electrical rhythmicity of the heart is disturbed, cardiac pumping is impaired or lost entirely, which can lead to fainting or sudden death; some 300,000 people die each year in the United States alone because of cardiac 'arrhythmias'. Several gene mutations predispose people to life-threatening arrhythmias, and, until now, all identified mutations have occurred in proteins that transport ions across cell membranes. On page 634 of this issue, Mohler and colleagues¹ show that a mutation in an 'adaptor' protein, which anchors ion transporters to specialized membrane domains, can also cause potentially fatal arrhythmias.

The events involved in generating heart beats are shown in Fig. 1a, overleaf. First, the pacemaker in the sinus node initiates electrical activity at a rate that is appropriate to the body's needs. Then the impulse is conducted throughout the heart, largely by means of

McFarland and Tang's device is based on majority carriers (electrons) only. As in dye-sensitized nanocrystalline solar cells⁴, this should render its performance less sensitive than conventional photovoltaic converters to impurities and imperfections. The fact that the device itself regenerates the dye layer, with no additional system required, is an advantage over other dye-sensitized semiconductors. But if the efficiency of this new converter is to be increased to practical levels, further development is needed, such as introducing light-trapping structures to boost the light-harvesting capacity of the molecular photoreceptors on the flat device surface.

Michael Grätzel is at the Institute of Photonics and Interfaces, École Polytechnique Fédérale, Ecublens, CH-1015 Lausanne, Switzerland.
e-mail: michael.gratzel@epfl.ch

1. McFarland, E. W. & Tang, J. *Nature* **421**, 616–618 (2003).
2. Gerischer, H. & Willig, F. *Top. Curr. Chem.* **61**, 31–84 (1976).
3. Scab, M. P. & Dersch, W. A. *Surf. Interf. Anal.* **1**, 2–11 (1979).
4. Grätzel, M. *Nature* **414**, 338–344 (2001).

the entry of Na⁺ ions into heart-muscle cells (myocytes) through specialized Na⁺-channel proteins. This depolarizes the cell (moving them away from their resting, negative intracellular charge), making Ca²⁺ enter the cell through Ca²⁺ channels and causing a temporary release of Ca²⁺ ions from an intracellular store called the sarcoplasmic reticulum. Contraction is thereby triggered. Once myocytes are depolarized, the Na⁺ channels become inactive, causing cells to enter a refractory, relatively inexcitable state until they return to their original potential by repolarization.

Repolarization is a delicate process, depending on an intricate balance between inward currents (consisting of Na⁺ or Ca²⁺ ions, which depolarize the cell interior) and outward currents (consisting of K⁺ ions, which move the cell back towards its resting potential). Any disruption to this balance can interfere with repolarization and predispose the heart to chaotic, potentially lethal arrhythmic activity. The 'QT' interval on an ECG recording corresponds to the time it takes for the ventricular chambers of the heart to repolarize, and a lengthened interval is a hallmark of repolarization defects.

A long QT interval is seen in several inherited human disorders, and Mohler *et al.*¹

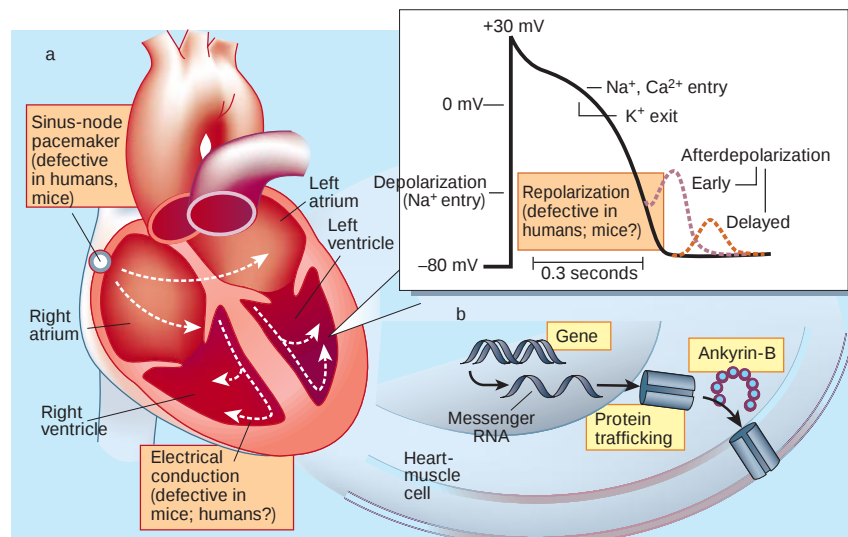


Figure 1 The electrical activity underlying heart beats, and the role of the ankyrin-B protein. **a**, Left, how electrical activity spreads through the heart; inset, the voltage inside heart-muscle cells. Processes or structures that are dysfunctional in mice that have just one functional copy of ankyrin-B or in human LQT4 patients (who have a mutation in ankyrin-B¹) are shown in orange. **b**, Processes and molecules involved in generating functional ion-transport proteins. From left to right, the gene encoding an ion transporter is transcribed into messenger RNA, which is used as a template for synthesis of the ion-transporter protein. This is then inserted into the membrane and held there with the help of ankyrin-B. Mutations in the ion-transporter gene, abnormalities in protein trafficking and defects in ankyrin-B¹ (yellow boxes) can all lead to cardiac arrhythmia and sudden death.

have studied the genetic basis of one of these — long-QT syndrome type 4 (LQT4). This unusual syndrome was described in a French family², and is characterized by a lengthened QT interval, potentially lethal arrhythmias, and disturbed sinus-node function. All cases of inherited long-QT syndrome analysed previously have involved mutations in ion-transport proteins, with the result that the outward K⁺ current is impaired or the inward Na⁺ current is enhanced. This reduces the net outward current, delaying repolarization and leading to arrhythmia.

In elegant work, Mohler *et al.* now show that LQT4 patients do not have mutations in an ion-transport protein. Instead, they have a single nucleotide change in the gene encoding the ankyrin-B protein. This mutation, whereby a guanine nucleotide in the gene replaces an adenine, results in one amino acid in the protein — the glutamic acid at position 1425 — being replaced by a glycine. The protein is therefore called E1425G ankyrin-B, as 'E' and 'G' represent glutamic acid and glycine respectively in the amino-acid code.

Supporting the importance of ankyrin-B defects to LQT4, Mohler *et al.* find that ankyrin-B heterozygous mice — which have only one functional ankyrin-B gene, rather than the usual two — have some features in common with LQT4 patients, including sinus-node abnormalities and arrhythmias. (There are also some differences: repolarization seems normal in the mice, and the animals have slowed cardiac electrical conduction, which has not been reported in patients.) Moreover, the authors show that

injecting normal ankyrin-B into myocytes from newborn heterozygous mice restores normal cell function, but injecting E1425G ankyrin-B has no effect.

How exactly does this cause arrhythmia? Ankyrin-B's fundamental role is to recognize certain other proteins and to ensure that they are inserted into appropriate domains of cell membranes^{3,4} (Fig. 1b). Several such proteins are involved in ion transport, including the pore-forming subunits of voltage-gated Na⁺ channels; the transmembrane Na⁺,Ca²⁺ exchanger; the Na⁺,K⁺-ATPase; and the Ca²⁺-release channel of the sarcoplasmic reticulum. Mohler *et al.* find that mutation of ankyrin-B disrupts the cellular localization of Na⁺,K⁺-ATPase and the Na⁺,Ca²⁺ exchanger. Moreover, myocytes from ankyrin-B heterozygous mice show defects in handling Ca²⁺ ions, which can cause afterdepolarizations (abnormal depolarizations that trigger arrhythmias; Fig. 1a).

This exciting study answers some questions, and raises others. For instance, the authors attribute LQT4 to abnormal Ca²⁺ handling by myocytes, resulting from abnormal Na⁺,K⁺-ATPase function. The Na⁺,K⁺-ATPase maintains Na⁺ and K⁺ gradients across cell membranes by pumping Na⁺ ions out of cells while bringing K⁺ ions in; so, if this protein functions abnormally (as occurs with an excess of the drug digitalis, for instance), Na⁺ ions accumulate inside cells. The Na⁺,Ca²⁺ exchanger then tries to compensate, exporting more Na⁺ ions and importing more Ca²⁺ ions. This leads to an increase in intracellular Ca²⁺



100 YEARS AGO

The rainfall of Madras has often been investigated as regards its relationship to the sun-spot curve, and the first indication of a probable periodicity with sun-spots was pointed out by Sir Norman Lockyer in 1872 and later by Dr. Hunter, in 1877... In a recent number of the United States *Monthly Weather Review*... Mr. M. B. Subha Rao, of the Madras Observatory, contributes an article on "The Rainfall in the City of Madras and the Frequency of Sun-spots."... Dealing with the variation of the rainfall and the sun-spot curve from the year 1811, he is led to deduce that the minimum rain "occurs almost exactly on the year of minimum frequency of sun-spots, the difference being only a year in a few cases." He finds, further, that the "maximum rainfall also takes place when we have the maximum frequency of sun-spots," but he guardedly adds that the difference amounts sometimes to two or three years. From *Nature* 5 February 1903.

50 YEARS AGO

Mr. H. Lloyd, of the Yorkshire Electricity Board, writes: "66,000 volts transmission lines in the Yorkshire and North Lincolnshire areas had been subject to a number of mysterious faults, which nearly always occurred about dawn, for several years before the trouble was found to be caused by kestrels which were escaping alive. The points of interest in pursuing the investigation were, first, how was it that the birds were surviving 38,000 volts applied across their bodies, and second, what was attracting them to the lines? The incidents always occurred in the early morning or in the misty weather when the birds' plumage could be assumed to be damp. The flashover is initiated across the damp wings when the bird is banking steeply with its back to the insulators, and current does not pass through the bird's body. The speed at which the bird is travelling takes it out of the path of the arc before it becomes too severely burnt... One day, a substation attendant, during misty weather, noticed a kestrel circling a string of insulators in ever-narrowing circles... The insulators, as is usual under humid conditions, were discharging and emitting an intermittent buzzing noise, and this seems to provide the attraction... This seems to imply that kestrels hunt by ear as well as sight." From *Nature* 7 February 1953.

stores, potentially promoting afterdepolarizations and arrhythmia^{5,6}.

However, alterations in Ca^{2+} handling that generate arrhythmias do not generally delay repolarization⁶. Yet abnormalities in repolarization are clearly seen² in LQT4 patients. In addition, the conduction slowing seen in ankyrin-B heterozygous mice cannot be attributed simply to Ca^{2+} -handling defects.

Another possibility is that the ankyrin-B mutation in LQT4 patients affects Na^+ channels. Cardiac myocytes from mice lacking both functional copies of ankyrin-B⁷ show defects in Na^+ channels: their ability both to carry Na^+ ions and to inactivate upon cellular depolarization is abnormal. Myocytes from ankyrin-B heterozygous mice show similar, if less severe, features⁷. Moreover, mutations in the pore-forming subunits of Na^+ channels that cause similar abnormalities produce conduction slowing, sinus-node dysfunction and repolarization defects⁸. Although the number and distribution of Na^+ channels in ankyrin-B heterozygous mice seem to be normal, their function might still be affected, as might the activity of other channels that control repolarization.

In summary, the work of Mohler *et al.*¹ reveals a new genetic cause for sudden cardiac

death, and provides insight into the crucial role of membrane-protein anchors. There is increasing evidence that proteins do not sit randomly in cell membranes, but require precise spatial localization. Mohler *et al.* have shown just how important this spatial organization can be. Although their study deals with a primarily cardiac disease, the ubiquitous distribution of ankyrin-B and related proteins raises the possibility that other tissues and organs are affected in this and similar genetic syndromes. It is also possible that non-genetic alterations in ion-channel function in cardiac diseases might be attributable to changes in ankyrin expression⁵. ■

Stanley Nattel is in the Department of Medicine and the Research Center, Montreal Heart Institute and University of Montreal, 5000 Belanger Street East, Montreal, Quebec H1T 1C8, Canada.
e-mail: nattel@icm.umontreal.ca

1. Mohler, P. J. *et al.* *Nature* 421, 634–639 (2003).
2. Schott, J. J. *et al.* *Am. J. Hum. Genet.* 57, 1114–1122 (1995).
3. Mohler, P. J. *et al.* *J. Cell Sci.* 115, 1565–1566 (2002).
4. Bennett, V. & Baines, A. J. *Physiol. Rev.* 81, 1353–1392 (2001).
5. Nattel, S., Khairy, P. & Schram, G. *Trends Cardiovasc. Med.* 11, 295–301 (2001).
6. Anderson, M. E. *J. Cardiovasc. Electrophysiol.* 13, 195–197 (2002).
7. Chauhan, V. S. *et al.* *Circ. Res.* 86, 441–447 (2000).
8. van de Berg, M. P. *et al.* *J. Cardiovasc. Electrophysiol.* 12, 630–636 (2001).

Earth science

Ins and outs on the ocean floor

John G. Sclater

Researchers investigating heat exchange between the ocean floor and the ocean have various long-standing issues to tackle. The latest measurements of heat flux in hydrothermally active crust will add to these debates.

Earth loses internal heat in two main ways: in cooling of the ocean lithosphere, originally generated at mid-ocean ridges, and in conduction across the boundary between the mantle and the continental and oceanic lithospheres. Heat exchange between the lithosphere and the oceans then occurs through both the hard but permeable crust at the top of the lithosphere, and the soft but impermeable sediments that cover much of the crust. At the axis of a mid-ocean ridge there are no sediments, and most heat loss is by hydrothermal discharge of sea water through the crust, in places through huge high-temperature geysers¹. Away from the axes, sediment starts to accumulate, and heat is lost by a combination of advection (water flow) through partially exposed 'basement' rocks, and possibly the sediments, and heat conduction through both rock and sediments. What, though, is the balance between these processes? And how is the hydrothermal outflow of water recharged?

Fisher *et al.*² (page 618 of this issue) show that, in one case at least, hydrothermal recharge and discharge occur entirely

through the basement rock, and that all the flux through the sediments is conductive. In studies of the Juan de Fuca ridge in the north-east Pacific, they find that an outcrop of permeable basement rock is the recharge point for a hydrothermal outlet more than 50 km away. This discovery adds the finishing touch to a theory originally put forward by Lister³, and should also reinvigorate research into conductive heat flux through the ocean floor.

In plate tectonics, the intrusion of hot material beneath a mid-ocean ridge creates oceanic lithosphere, which behaves like a plate. As it moves away from the ridge, the lithosphere cools, contracts and is covered in sediment, before finally disappearing into the mantle beneath an oceanic trench. In general, the sedimentary layer thickens with increasing age of the ocean floor, and outcrops of basement rock become sparser before being completely covered by sediment on floor that is more than 60–100 million years old (Fig. 1). According to Lister³, warm circulating water escapes into the ocean through rock outcrops and the recharge area is a similar outcrop nearby. In an area totally covered by

sediment, he believed, hydrothermal circulation occurs within the basement rock but there is no connection to the sea water above.

Geochemical measurements on pore fluids show that thin layers of sediment do permit the passage of upwelling and downwelling fluids⁴. Thus, an alternative view was that, in sedimented regions, recharge occurred through the flow of sea water downwards, through sediments close to the region of discharge. Detailed heat-flux surveys did not always reveal the low and high values in close proximity, however, and sea-water flow through the thick, impermeable sediments that cover most of the ocean floor seemed unlikely. Fisher *et al.*² now show, in an area almost completely covered by sediment, that outcrops of basement rock can indeed be the source of the return flow. They carried out heat-flux surveys both around and between two outcrops, 50 km apart, in the sediment-filled basin east of the Juan de Fuca ridge. The measurements indicate vigorous discharge at one of them (a narrow outcrop known as Baby Bare), and a substantial recharge through the much larger one (Grizzly Bare), with no evidence for water flow through the sediments in this area.

The bigger picture here concerns various models of the lithosphere. The match between conductive heat flux and age, and subsidence of the mid-ocean ridge with age, initially led to a model of the lithosphere as a flat plate about 125 km thick⁵. Given the lack of a physical justification for that, an alternative proposal was to view the lithosphere as a cooling, rigid, thermal boundary layer that thickens at a rate that is proportional to the square root of time⁶ (like ice on a still lake when the temperature is well below freezing). But this model does not account for the 'flattening' of ocean-floor depth with age and predicts a heat flux through the old ocean basins that lies well below that observed. Another suggestion is that oceanic lithosphere consists of a rigid mechanical layer on top of a lower-viscosity, thermal boundary layer⁷, with the argument that if both layers increase in thickness with age, the lower one will become unstable after a specific time: this mechanism would have the same effect on the heat-flux and depth measurements as the plate model, and provides a physical justification for it.

Most recently, Nagihara, Lister and I went back to detailed surveys of thermal flux and isostatically corrected depth measurements to show that the data lie on the expected curve for a boundary-layer model, except for being offset in age⁸. This we attribute to varying amounts of 'thermal rejuvenation' of the lithosphere. It is unclear how well this observational approach, relying on the old heat-flow data, is compatible with a two-layer structure for the lithosphere.

Given that the lithosphere marks the upper boundary for convection in Earth's mantle, a quantitative grasp of its behaviour is

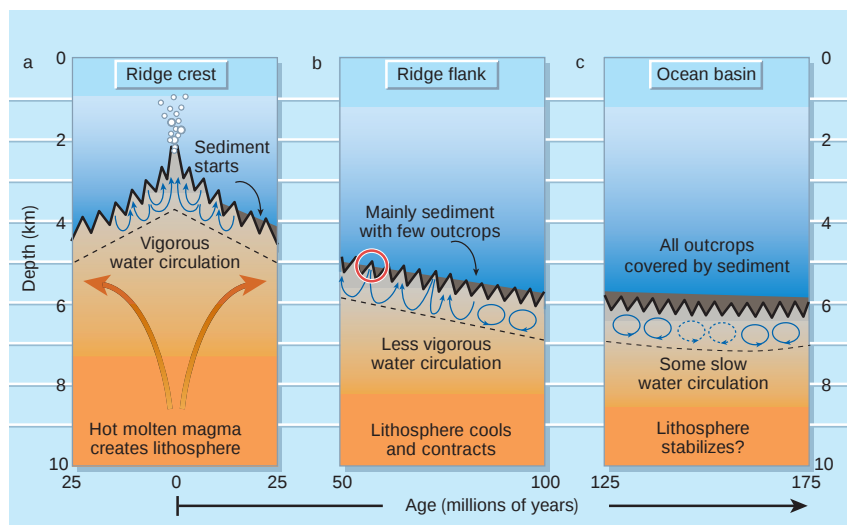


Figure 1 Changes in ocean-floor depth, sediment thickness and hydrothermal circulation with increasing age of the lithosphere. The ocean floor is depicted in black, sediment in brown and hydrothermal circulation in blue. In a, heat loss is by advection; in b, by advection and conduction; and in c by conduction alone. Fisher *et al.*² deal with younger crust than that shown in b, but this panel illustrates the flow pattern for the flank of the Juan de Fuca ridge that they surveyed. They find that hydrothermal recharge is occurring through an outcrop of basement rock (red circle) 50 km distant from the point of discharge, and that any heat flow through sediments is by conduction, not advection. (Diagram based on ref. 3.)

essential for understanding mantle dynamics. A difficulty, however, is that the scatter in the ocean-floor subsidence data increases dramatically for ocean floor more than 80 million years old⁹. The scatter in the conductive heat-flux values is reduced for this age of crust, but lack of knowledge of seawater flow in sedimentary areas with basement outcrops has made it difficult to have confidence in these values. Importantly, then, Fisher *et al.*² show that the heat-flux values lie close to those predicted for the age of crust between the outcrops. This implies that there is little deficit in the conductive heat loss, and thus that the flow of sea water must be slow enough to have little or no effect on the conductive flux.

This result, and discovery that the discharge and recharge occur at outcrops, together have major implications for the use of heat-flux measurements to study how the lithosphere changes with age. They explain why most detailed studies on ocean floor more than 60 million years old have little scatter in the values¹⁰. Relatively thick sedimentary layers cover most ocean crust older than that, and either no heat is lost by advection because there are no outcrops or, if there are, the flow is so slow that it has only a small effect on the conductive flux (Fig. 1).

This is an exciting time to be studying the heat regime of the lithosphere. Now that we are beginning to understand patterns of hydrothermal circulation in the basement rock, it is clear that heat-flux surveys over broad areas well covered by sediment will provide reliable estimates for the conductive flux. Moreover, direct measurements of the density structure of the upper mantle are

emerging. Variations in the group velocity of surface waves as they cross the Pacific provide evidence¹¹ that, on crust 70–100 million years old, the thickening of the Pacific lithosphere flattens relative to that predicted by the

cooling boundary-layer model — so constituting evidence for thermal rejuvenation of the lithosphere as it ages.

At a meeting in September last year, the installation of leapfrogging arrays of ocean-bottom seismometers was proposed, with the aim of further refining models of the lithosphere and coupling them to convection patterns in the mantle¹². Combining the seismic information with new data on heat flux, ocean-floor depth and the gravity field should greatly advance our understanding of global circulation in the mantle.

John G. Sclater is in the Geosciences Research Division 0215, Scripps Institution of Oceanography, University of California San Diego, La Jolla, California 92093-0220, USA.

e-mail: jsclater@ucsd.edu

1. Van Damm, K. *Nature* **412**, 127–128 (2001).
2. Fisher, A. T. *et al.* *Nature* **421**, 618–621 (2003).
3. Lister, C. R. B. *Geophys. J. R. Astron. Soc.* **26**, 515–535 (1972).
4. Kastner, M. & Rudnicki, M. D. in *Hydrology of the Oceanic Lithosphere* (eds Davis, E. & Elderfield, H.) (Cambridge Univ. Press, in the press).
5. Parsons, B. & Sclater, J. G. *J. Geophys. Res.* **82**, 803–827 (1977).
6. Crough, S. T. *Geophys. J. R. Astron. Soc.* **55**, 451–470 (1978).
7. Parsons, B. & McKenzie, D. P. *J. Geophys. Res.* **83**, 4485–4496 (1978).
8. Nagihara, S., Lister, C. R. B. & Sclater, J. G. *Earth Planet. Sci. Lett.* **39**, 91–104 (1996).
9. Stein, C. A. & Stein, S. *Nature* **359**, 123–129 (1992).
10. Von Herzen, R. P. in *Hydrology of the Oceanic Lithosphere* (eds Davis, E. & Elderfield, H.) (Cambridge Univ. Press, in the press).
11. Ritzwoller, M. H., Shapiro, N. M., Landuyt, M. & Zhong, S. *Eos* **83**, S345 (2002).
12. Ocean Mantle Dynamics Implementation Plan Report of a Community Workshop (National Science Foundation, Washington DC, 2003).

Ecology

Shady deals with lichens

Peter D. Moore

In tropical forests, many species of lichen gain a place in the sun by living on the leaves of other plants. It seems they can do so without compromising the photosynthetic capacity of their host.

Life can be challenging for a small plant in a tropical rainforest. Most plants of lesser stature solve the problem of light acquisition by being epiphytic — they use taller plants as support systems to gain a place in the sun. Epiphytes include a wide range of plants, from liverworts to orchids, and some, known as epiphyllous species, even adhere to leaf surfaces. Most epiphytes present no direct threat to the host species, although supporting them may impose a burden and, in the case of epiphylls, could interfere with light availability to the host leaves.

Lichens are one of the most common groups of epiphyllous organism. Writing in *Functional Ecology*, Anthony, Holtum and Jackes¹ describe how they assessed the potential shading impact of lichens on their host. The authors show that the host leaves are photosynthetically resilient, and can com-

pensate for the shading effect of lichens, maintaining their overall productivity despite the presence of their uninvited guests.

Epiphylls are abundant in tropical rainforests² and have been generally assumed to have an adverse effect on the photosynthetic capacity of a leaf — they both reflect and absorb light, leaving the host leaf in increased shade. Working on a coffee plantation in Mexico, Roskoski³ found up to 20% of each leaf was covered by epiphylls, but surprisingly, the overall productivity of the crop did not seem to be affected. Studies⁴ in the rainforests of Panama, however, revealed that epiphylls reduced the amount of light received by leaves by 55–85%, with an estimated 20% reduction in overall photosynthesis. The precise impact of epiphylls on their host's productivity is thus a matter of debate.

Intuitively, one might suppose that such

an impediment to light supply would be a disadvantage for the host plant, but the effects of lichen cover on leaf surfaces have never been quantified. Affected leaves might be able to adjust to the new conditions of shading, by biochemical means and by structural alterations, such as the movement of chloroplasts to new locations within the cell⁵. For most tropical rainforest leaves, the lichen coverage increases gradually as they age. So leaves might be able to adapt to the new light conditions and retain their productive potential.

Anthony and colleagues¹ have studied lichens on the leaves of the palm *Calamus australis* in northern Queensland, Australia. Their first concern was to determine precisely how light conditions alter for a host leaf as a consequence of developing lichen cover. They found that light interception values for lichens had a mean of 48%, so the lichens effectively halved the amount of light reaching the host leaves. But did the lost light lie within the wavelengths required for photosynthesis by the host? It did — the light absorbed by the lichens had a similar spectral composition to that absorbed by the host leaf.

However, the possibility remained that compensation could occur by adjustment of a leaf's photosynthetic systems. Anthony *et al.* investigated this by examining the relationship between light intensity and carbon gain by photosynthesis in parts of the leaf either covered by or clear of lichens. They found that the two areas behaved quite differently. The part covered by lichens showed characteristics of a classic 'shade leaf', having a capacity to fix carbon effectively even at very low levels of irradiance. Respiration rates and the light compensation point (the irradiance level at which carbon fixation balances respiratory demand) remained unaltered, but the more efficient use of low light levels showed that lichen-covered parts of the leaf had acclimated to the new conditions. Checking the chlorophyll levels in different parts of the leaves revealed that the shaded portions had increased concentrations, although the ratio of the two main types of chlorophyll (*a/b*) remained the same. The gradual accretion of lichen cover during the lifetime of a leaf evidently allows this adjustment to take place, and local photosynthetic efficiency can thereby be maintained. Modelling the total carbon gain for entire leaves, Anthony *et al.* found that the general carbon economy of the host was not greatly affected by the presence of the epiphylls.

These findings would appear to resolve the physiological conundrum, first described for coffee plants³, in which increased epiphyte shading resulted in no loss of productivity. But other questions remain, such as those relating to coevolution of organisms within rainforest and the consequent high biodiversity of this habitat. The leaf can

adapt to the lichen epiphyte, but has the lichen also adapted in a way that will ensure its host's survival? For example, the fact that the lichen steals only half of the available light could be the result of a trade-off between further productivity by the epiphyll and the consequent danger of reducing the productivity of the host leaf and possibly destroying it. Then there is the question of why leaves with a particularly long life-span generally have low levels of epiphylls⁴. Is the development of some chemical or structural defence mechanism by these highly persis-

tent leaves an alternative strategy to that of appeasement? ■

Peter D. Moore is in the Department of Life Sciences, King's College London, London SE1 9NN, UK.

e-mail: peter.moore@kcl.ac.uk

1. Anthony, P. A., Holtum, J. A. M. & Jackes, B. R. *Funct. Ecol.* 16, 808–816 (2002).
2. Richards, P. W. *The Tropical Rainforest* (Cambridge Univ. Press, 1957).
3. Roskoski, J. P. *Oikos* 37, 252–256 (1981).
4. Coley, P. D., Kursar, T. A. & Machado, J.-L. *Ecology* 74, 619–623 (1993).
5. Kasahara, M. *et al. Nature* 420, 829–832 (2002).

Attophysics

Ultrafast control

Philip H. Bucksbaum

The successes of two pioneering groups are now brought together to create trains of identical ultrashort laser pulses that can control what's happening inside an atom.

One attosecond is one-millionth of one-millionth of one-millionth of a second. If the next second of your life were expanded to fill the history of the Universe, then one attosecond would correspond to less than one second of that history. In one attosecond (10^{-18} s), light travels slightly further than the length of a single water molecule, and the molecule itself appears utterly frozen. The natural vibrations of a molecule take place over tens of thousands of attoseconds, and rotations take millions of attoseconds — a relative eternity. Even the rapid motion of atomic electrons is typically measured in hundreds to thousands of attoseconds; and a single oscillation of a wave of visible light lasts nearly 2,000 attoseconds.

The title, then, of the report by Baltuška *et al.*¹ on page 611 of this issue — "Attosecond control of electronic processes by intense light fields" — is impressive indeed. The authors claim to have harnessed light on an attosecond scale, and used the light to control electrons with similar precision. This

work combines the efforts of two creative research groups in optical science: the precision measurements of Theodor Hänsch's group at the Max Planck Institute for Quantum Optics, and the ultrafast-laser expertise of Ferenc Krausz's group at the Technical University of Vienna. Krausz, Hänsch and their colleagues have achieved this dizzying feat by delicately stacking together two techniques that have grown out of their research into ultrafast-laser technology: the production of subfemtosecond soft X-ray pulses², and the production of an all-optical-frequency standard³.

Both of these achievements are based on the Kerr-lens mode-locked laser, the workhorse of ultrafast science, invented only a dozen years ago and now found in most of the world's ultrafast-research laboratories⁴. It produces a continuous train of 10–100-femtosecond pulses (10,000–100,000 attoseconds) in the near-infrared region of the optical spectrum, through a fortuitous nonlinear effect known as the Kerr lens (named

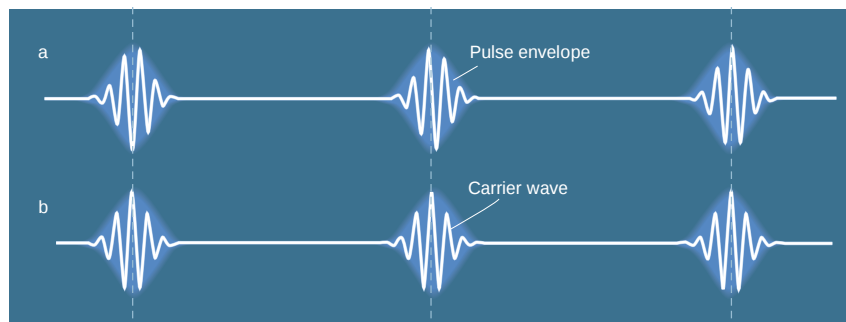


Figure 1 Locking onto the laser frequency. **a**, The output of a typical ultrafast laser, based on the Kerr lens, forms a train of non-identical pulses: the optical carrier wave inside each pulse is shifted with respect to the pulse envelope, because its frequency is not an integer multiple of the pulse repetition frequency. **b**, If, however, the laser frequency is stabilized (using an optical comb generator³), truly identical pulses are created.

after the Scottish physicist John Kerr, 1824–1907). The lens is created by the light itself, and forces the photons in a laser cavity to bunch up into extremely short pulses. These pulses can be thought of as electromagnetic waves confined in a single short envelope, bouncing back and forth between the end mirrors of the laser cavity. If one of the end mirrors is partly transmitting, a small portion of the pulse leaks out on each round trip, creating a train of output pulses.

But, although transmitted through the mirror on successive round trips, the pulses in the train do not look identical. The optical carrier wave appears shifted by some fraction of a cycle on every bounce (Fig. 1a). This is because the carrier frequency is not commensurate with the pulse repetition frequency — that is, the ratio of the optical frequency to the repetition frequency is not an integer. If the ratio were an integer, then the optical frequency could be simply related to the round-trip frequency.

Several years ago, Hänsch realized that if this defect could be overcome, it would be possible to build all-optical-frequency standards, which would significantly simplify and improve the precision of time measurement. His group found a way to do this, and the result, known as the ‘optical comb generator’, is revolutionizing the time standard. The details involve ‘holy fibres’, self-phase modulation and second-harmonic generation, a fascinating story which has been told in a previous issue of *Nature*³. Suffice to say here that Hänsch learned how to make a Kerr-lens laser in which every pulse in the train is truly identical (Fig. 1b).

Meanwhile, Krausz’s group was pursuing a different application of Kerr-lens ultrafast lasers. They amplified light pulses only a few cycles long to peak powers of 100 billion watts, to create bursts of short-wavelength radiation. The generation of energetic ultra-short pulses with these lasers is not trivial. Straightforward laser amplification will destroy both the laser and its light because the intensities are high enough to rip atoms apart. This catastrophe can be averted using exotic techniques with equally exotic names: chirped-pulse amplification, hollow-fibre self-phase modulation and chirped mirror compression. This is ultrafast optics raised to a high art, and Krausz is a virtuoso.

When the high-powered light is focused into atoms in a vapour, the atoms react in a number of ways. Some of them ionize rapidly, creating a shower of energetic electrons through what is called ‘above-threshold ionization’⁵. Other atoms absorb laser light and re-radiate at very high harmonics (or multiples) of the laser frequency⁶. This high-harmonic radiation constitutes a unique source of ‘soft’ X-rays, with pulse durations that can be less than a femtosecond².

Krausz’s group and others have studied the mechanism for high-harmonic genera-

tion. Theories suggest that the high harmonics are due to coherent radiation created by the recollisions of ionized electrons with their parent atoms^{7,8}. If so (and experiments seem to support this), then the soft X-rays should arrive in attosecond bursts, occurring on every cycle of the laser field. Thus, the pattern of X-ray bursts matches the pattern of the optical carrier wave within the pulse envelope — and this is where the work of Hänsch’s group and of Krausz’s group now comes together¹.

When Krausz’s laser is stabilized by Hänsch’s technique, each 100-billion-watt pulse from the amplifier looks identical. This means that electrons ripped from the atoms in the focus of the laser beam follow identical paths, recollide with their parent atoms at identical stages of the pulse, and produce optical and electron spectra that record this violent motion. All previous experiments have been forced to average over non-identical laser pulses, and so many special characteristics of the atomic-scale process were washed out. In particular, the harmonics spectrum is now revealed to be not always ‘harmonic’ (an integer multiple of the pump frequency) after all. Rather, the spectrum is modulated by the coherent interference between successive X-ray bursts initiated at

different parts of successive cycles of the driving light field.

The phase-locking to produce identical pulses is successful down to a small fraction of an optical period, amounting to a temporal smearing of less than 200 attoseconds. Now, 200 attoseconds is not one attosecond, so the title of Baltuška and colleagues’ paper¹ is perhaps rushing things a bit. Nonetheless, this is a remarkable achievement. This level of precision can capture the motion of photo-excited electrons as they move around their parent ion. The success of Hänsch, Krausz and their colleagues in creating frequency-locked high-field laser pulses marks the beginning of the era of ‘attophysics’ — the study of physical processes on the attosecond timescale. ■

Philip H. Bucksbaum is at the FOCUS Center, University of Michigan, Ann Arbor, Michigan 48109-1120, USA.

e-mail: phb@umich.edu

1. Baltuška, A. *et al.* *Nature* 421, 611–615 (2003).
2. Hentschel, M. *et al.* *Nature* 414, 509–513 (2001).
3. Udem, T., Holzwarth, R. & Hänsch, T. W. *Nature* 416, 233–237 (2002).
4. Spence, D. E., Kean, P. N. & Sibbett, W. *Opt. Lett.* 16, 42–44 (1991).
5. Agostini, P., Fabre, F., Mainfray, G., Petite, G. & Rahman, N. K. *Phys. Rev. Lett.* 42, 1127–1130 (1979).
6. Ferray, M. *et al.* *J. Phys. B* 21, L31–L35 (1988).
7. Corkum, P. B. *Phys. Rev. Lett.* 71, 1194–1197 (1993).
8. Krause, J. L., Schafer, K. J. & Kulander, K. C. *Phys. Rev. Lett.* 68, 3535–3538 (1992).

Cancer

More than skin deep

Diana Bolotin and Elaine Fuchs

Activation of the nuclear factor NF- κ B has been linked with many human cancers. But in tumours of some tissues this molecule is inactive. New work examines how turning off NF- κ B promotes cancer in one tissue, the skin.

What do cell death, cell proliferation, immune responses and tumour development have in common? The answer is that they are just a few of the many biological processes in mammals that are regulated by proteins of the nuclear factor NF- κ B family^{1,2}. Widely expressed throughout an organism, the NF- κ B subunits are also exquisitely regulated in terms of their ability to enter the cell nucleus and activate key genes. However, although much is known about these proteins, many issues remain unexplored. For instance, studies of the immune system and several cell lines have indicated a role for NF- κ B in promoting cell proliferation and protecting against programmed cell death (apoptosis)^{2,3}. So it is puzzling that, in the skin, these proteins seem to oppose proliferation^{4,5}. On page 639 of this issue⁶, Dajee *et al.* address the paradox that inhibition of NF- κ B in the skin can lead to increased cell proliferation — and cancer. Their findings may be important not only for understanding how skin tumours develop, but also for exploring

the design and use of NF- κ B inhibitors in treating cancer.

Our cells regulate the activity of NF- κ B proteins — which exist as dimers — largely by controlling their intracellular localization, using the I κ B molecule⁷. When bound to I κ B, NF- κ B proteins are excluded from the nucleus and so are prevented from activating their target genes. However, certain stimuli can induce an enzyme complex, the I κ B kinase, to modify I κ B, targeting it for degradation. The NF- κ B dimer is thereby released and moves into the nucleus, resulting in the expression of genes with functions appropriate to the stimuli.

For many years, the function of NF- κ B in the organism seemed relatively straightforward, fitting a profile for the protein as the cell’s protector against apoptosis and an enhancer of proliferation. So, in several organs and tissues, high levels of NF- κ B can prevent apoptosis by activating genes for anti-apoptotic proteins such as Bcl-xL and cellular inhibitors of apoptosis². And in the immune system, where these proteins have

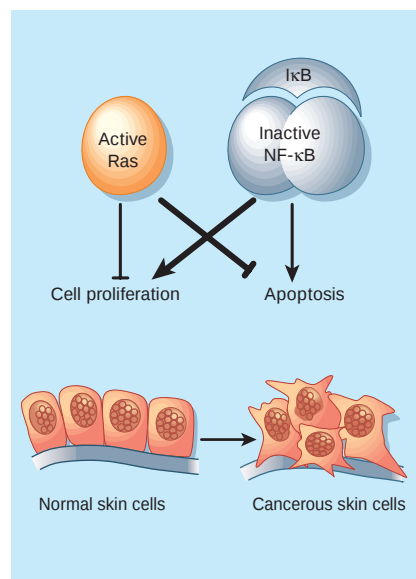


Figure 1 Molecular interactions that can lead to skin cancer. Dajee *et al.*⁶ have investigated what happens when NF- κ B is inhibited in the skin in the context of activated Ras protein. To do this, they inserted active Ras and a stable I κ B protein (which inhibits NF- κ B) into human skin keratinocytes, and then grafted these cells onto mice. The result is that inhibition of NF- κ B overrides the proliferation arrest induced by active Ras; and the active Ras overrides the apoptosis induced by inhibited NF- κ B. This leads to the development of cancers resembling human squamous cell carcinomas (SCCs). Other molecular features of these cancers are the maintenance of telomeres (protective caps at the end of chromosomes), and a dependence on the proteins laminin 5 and integrin α 6 β 4.

been extensively studied, high NF- κ B levels also stimulate the expression of growth-factor proteins and regulators of the cell-division cycle, often promoting cell proliferation². The dual effects of stimulating cell proliferation and preventing cell death would seem to form a potent potion for uncontrolled cell division, and possibly cancer. Indeed, sustained activation of NF- κ B has been linked to several human cancers².

Although many of the data on NF- κ B fit neatly into this model, there is increasing evidence that matters are not so straightforward. Most recently, for instance, some colon-cancer cells have been found to have decreased NF- κ B activity, through an interaction with the β -catenin protein⁸ (often deregulated in colon cancers). Curiously, earlier data had shown that sustained NF- κ B activity prevents the proliferation of normal skin epidermal cells⁴. And in mice genetically engineered to overexpress a stabilized I κ B in the skin epidermis, NF- κ B is permanently inhibited and the skin becomes sensitized to developing squamous cell carcinomas (SCCs)^{4,5}. Dajee *et al.*⁶ now show that, in

human skin, inhibiting NF- κ B while activating the Ras protein leads to cell proliferation and cancer.

Arguably the most well-known tumour-causing molecule, Ras has been central to cancer research because of its ability to cause uncontrolled cell proliferation, to modulate apoptosis and to stimulate tumour spread (metastasis) — all hallmarks of cancer development^{9,10}. As Dajee *et al.* show, however, matters are again different in human skin: activating Ras leads to proliferation arrest as cells switch on appropriate counteracting mechanisms. Dajee *et al.* further find that activation of Ras increases NF- κ B activity, suggesting a possible mechanism underlying Ras's inhibitory effects on cell proliferation. Therefore, the authors decided to examine the interplay between Ras activation and NF- κ B inhibition in the development of human skin cancer.

To relate their findings to human cancer, Dajee *et al.* first used a retrovirus to insert stabilized I κ B and activated Ras into human keratinocytes (skin cells), and then grafted these cells onto immune-deficient mice. Intriguingly, the resulting inhibition of NF- κ B activity synergized with the activated Ras to generate skin tumours resembling human SCCs. It seems that inhibiting NF- κ B allowed keratinocytes to escape the Ras-induced proliferation arrest and to progress through the cell cycle. And, in an unfortunate molecular symbiosis, active Ras rendered these cells resistant to death induced by NF- κ B inhibition, further favouring uncontrolled proliferation (Fig. 1).

The potential of a tumour to invade tissues often correlates with its ability to express certain classes of integrin receptor — dimeric cell-surface proteins that interact with the extracellular matrix and trigger signalling cascades that can affect cell proliferation and migration^{11,12}. The expression of integrin α 6 β 4 has been associated with the propensity of SCC cells to invade and spread¹³. Dajee *et al.* now show that the SCCs generated by stabilized I κ B and activated Ras express high levels of both integrin α 6 β 4 and laminin 5, its partner in the extracellular matrix.

The authors then take their analysis a step further, and demonstrate the functional importance of elevated levels of integrin α 6 β 4 in their model of human SCC. The analysis is particularly interesting in that Dajee and colleagues' approach enabled them to insert stabilized I κ B and activated Ras into keratinocytes from patients with junctional epidermolysis bullosa. This blistering disease involves mutations in integrin α 6, integrin β 4 or laminin 5. In this case no tumours were formed, underscoring the importance of laminin 5, and integrin β 4 in particular, in SCC development.

A number of intriguing questions remain unanswered. For instance, the exact mecha-

nism underlying the interplay between NF- κ B inhibition, integrins and Ras in the development of skin cancer requires further study. Furthermore, it is still unclear why in some tissues NF- κ B activation leads to tumour development, whereas in skin it seems to lead to proliferation arrest, protecting against cancer. The newly found ability of β -catenin to inhibit NF- κ B⁸ also raises the question of whether this interaction, thus far described only in colon-cancer cells, might also regulate NF- κ B in normal cells. Whatever the answers, it seems clear from several recent studies that NF- κ B regulation is vital for cells to orchestrate the balance between cell death, proliferation and differentiation. Our understanding of the effects of NF- κ B activation or repression is likely to require elucidation of which genes in a particular cell are activated by NF- κ B, and the status of NF- κ B's partners, which may include β -catenin as well as I κ B.

These complexities raise caveats about the use of NF- κ B inhibitors to treat cancer^{14,15}. Many naturally occurring and synthetic compounds that block NF- κ B activation have been identified, and have been suggested for use in cancer treatment in conjunction with standard chemotherapy. In the skin and colon, however, compounds that bring about sustained NF- κ B inhibition might instead promote tumours, particularly when Ras is activated. These new twists will certainly provide fresh challenges for the design of NF- κ B inhibitors in cancer therapy. ■

Diana Bolotin and Elaine Fuchs are at the Rockefeller University, Howard Hughes Medical Institute, 1230 York Avenue, Box 300, New York, New York 10021-6399, USA.

e-mail: fuchs@mail.rockefeller.edu

1. Sen, R. & Baltimore, D. *Cell* 47, 921–928 (1986).
2. Karin, M., Cao, Y., Greten, F. R. & Li, Z. W. *Nature Rev. Cancer* 2, 301–310 (2002).
3. Mayo, M. W. *et al. Science* 278, 1812–1815 (2002).
4. Seitz, C. S., Lin, Q., Deng, H. & Khavari, P. A. *Proc. Natl Acad. Sci. USA* 95, 2307–2312 (1998).
5. van Hogerlinden, M., Rozell, B. L., Ahrlund-Richter, L. & Toftgard, R. *Cancer Res.* 59, 3299–3303 (1999).
6. Dajee, M. *et al. Nature* 421, 639–643 (2003).
7. Ghosh, S. & Karin, M. *Cell* 109, S81–S96 (2002).
8. Deng, J. *et al. Cancer Cell* 2, 323–334 (2002).
9. Hanahan, D. & Weinberg, R. A. *Cell* 100, 57–70 (2000).
10. Macaluso, M. *et al. J. Cell. Physiol.* 192, 125–130 (2002).
11. Watt, F. M. *EMBO J.* 21, 3919–3926 (2002).
12. Giancotti, F. G. & Ruoslahti, E. *Science* 285, 1028–1032 (1999).
13. Tennenbaum, T. *Cancer Res.* 53, 4803–4810 (1993).
14. Darnell, J. E. Jr *Nature Rev. Cancer* 2, 740–749 (2002).
15. Bharti, A. C. & Aggarwal, B. B. *Biochem. Pharmacol.* 64, 883–888 (2002).

Correction

In Richard O. Prum's News and Views article, "Dinosaurs take to the air" (*Nature* 421, 323–324; 2003), an incorrect version of Fig. 1 was printed. The correct version, in which the label "Theropod dinosaurs" extends to take in "Archaeopteryx" and "Other birds", has appeared on *Nature's* website since publication.

Obituary

Grote Reber (1911–2002)

On 5 May 1933, *The New York Times* carried a front-page article reporting Karl Jansky's discovery of radio noise emanating from the centre of the Galaxy. Jansky had been investigating the source of interference in the short-wave radio links that were then used for transatlantic telecommunications. But although his discovery received considerable popular attention, it made no impact on mainstream astronomy. As Grote Reber commented, "The astronomers of the time didn't know anything about radio or electronics, and the radio engineers didn't know anything about astronomy".

Reber was then a 22-year-old engineering graduate of the Armour Institute of Technology in Illinois with a specialty in electronics and communications. He had built and operated his own amateur radio station, W9GFZ, and was looking for new challenges when he heard Jansky's 'star noise' as it was rebroadcast by NBC radio. Reber tried to contact professional astronomers about it, but they showed little interest. "So," he later related, "I consulted with myself and decided to build a dish."

With \$2,000 of his own money — about equivalent to his annual salary, but which he claimed he managed to afford by using public transport instead of buying a car — he built a 32-foot antenna in his mother's backyard. He took astronomy courses at the University of Chicago, and using his experience and skills as an electrical engineer and radio amateur he designed, built and tested a series of sensitive radio receivers. After many failures, in the spring of 1939 Reber finally succeeded in detecting Jansky's galactic radio noise.

Automobile-ignition noise interfered with Reber's observations, so he observed at night, laboriously writing down the readings from his detector every minute. By day, he returned to his job in Chicago, catching a few hours' sleep each evening before returning to his night's observations. At weekends, he analysed his data, and eventually produced the first high-resolution radio map of the sky. He also discovered the surprisingly intense radio-noise storms from the Sun.

At first, his work was received with scepticism by the astronomy community, and he had great difficulty getting his papers accepted for publication in the astronomical literature. According to Reber, the professionals thought "the whole affair was at best a mistake and at worst, a hoax".



Father of radio astronomy

But Reber's radio map provided the incentive for the dramatic growth in radio astronomy that occurred after the end of the Second World War. Former radar scientists and astronomers, primarily in the United Kingdom, Australia and the Netherlands, built a series of ever more powerful radio telescopes. With them, they made remarkable discoveries that have changed our fundamental understanding of the Universe. The enormously energetic clouds of relativistic electrons and cosmic jets that extend up to millions of light years into space, as well as quasars, pulsars, gravitational lensing, cosmic microwave masers, extrasolar planetary systems, complex interstellar molecules, and the cosmic microwave background radiation were all discovered by radio telescopes. Radio telescopes have also been used to measure the relativistic bending of electromagnetic waves that pass near the Sun, and to demonstrate the existence of gravitational radiation and measure continental drift.

Since Reber's early work, radio astronomers have steadily moved to ever shorter wavelengths in the quest for better angular resolution and for the multitude of molecular transitions that exist in the millimetre and submillimetre wavebands. Reber, however, as usual departing from conventional 'wisdom', concentrated in his later research on extremely long wavelengths. Working with Bill Ellis in

Tasmania, where the ionospheric attenuation of radio waves is at a minimum, Reber designed and built a number of antennas to study galactic radio emission at wavelengths of a few hundred metres, as well as building an energy-efficient home in Bothwell, where he lived for many years.

Except for a brief period between 1948 and 1951 when he was at the National Bureau of Standards in Washington, Reber worked as an amateur whose research was directed only by his curiosity and imagination. He paid no attention to establishment science — except to express his disregard. "There were no self-appointed pontiffs looking over my shoulder giving bad advice," he explained. "The kinds of things I want to do are the kind establishment men will not have any part of."

In addition to his pioneering work in radio astronomy, Reber also published research in a variety of fields ranging from radio circuitry and ionospheric physics to the growth of beans and the carbon dating of aboriginal campfire sites. Throughout his career, he was unable to secure funding from any of the conventional sources, such as the US National Science Foundation or the Department of Defense. Instead, he relied on modest support from the New York-based Research Corporation, as well as his own personal funds.

Grote Reber was the world's first radio astronomer. His 32-foot backyard telescope was the forerunner of the large steerable radio telescopes of today, as well as the smaller satellite-television dishes found in so many homes. Although Karl Jansky was the first to detect cosmic radio emission, it was Reber who, through his innovative experiments, forceful personality and stubborn persistence, finally convinced astronomers that it might be important and opened a new window on the Universe.

Until a few months before his death on 20 December 2002, two days before his 91st birthday, Reber continued to be active on a variety of scientific, political and social issues. He argued, with equal enthusiasm, against the Big Bang Universe and the increasing use of fossil fuels, and took a public stand against 'big science'. He was described by some as a genius, by others as a crackpot; but he was ultimately recognized by the astronomy community with the award of most of its major prizes.

K. I. Kellermann

K. I. Kellermann is at the National Radio Astronomy Observatory, 520 Edgemont Road, Charlottesville, Virginia 22903-2475, USA. e-mail: kkellerm@nrao.edu

Molecular biology

Dynamics of condensed DNA

Science **299**, 719–721, 721–725 (2003)

Inside the cell nucleus, DNA is packaged into a compact form called chromatin. Of the two main types of chromatin, heterochromatin is particularly highly condensed and comprises genetic information that is generally kept silent. The protein HP1 is thought to crosslink chromatin in a stable fashion, creating a dense heterochromatin environment that is impermeable to gene activators.

Two groups now challenge this view. Using cells that produce HP1 fused to green fluorescent protein (GFP), they determined HP1's mobility by measuring fluorescence recovery after bleaching. Richard Festenstein *et al.* generated mice that express GFP–HP1 in T cells, whereas Thierry Cheutin *et al.* looked at Chinese hamster ovary cells. Fluorescence recovery was rapid, indicating that HP1 is highly mobile. There was an immobile fraction in resting T cells, but in activated T cells the protein's mobility increased, and the immobile fraction was smaller.

So it seems that HP1 can be released from and then re-bind to heterochromatin. The authors conclude that heterochromatin is not, as previously suspected, inaccessible to other proteins: the continuous exchange of HP1 may allow gene regulators to compete for binding.

Arianne Heinrichs

Palaeontology

Going to work on an egg

Geology **31**, 39–42 (2003)

Among the records of early life are tiny fossilized eggs and embryos, presumably produced by invertebrate animals and dating to around 570 million years ago. Derek Martin and colleagues have carried out a lab study of how such fossils might have formed, using eggs of the European lobster (*Homarus gammarus*).

The eggs were kept in vials of artificial sea water, with a soupçon of added sediment, simulating coastal or estuarine conditions. After 36 days, the eggs were examined by scanning electron microscopy. Mineral coatings some 50 nm thick were evident, with preservation of surface features seen in both fossil and fresh eggs (see picture), and the process occurred in the absence of larger organisms as a source of mineralizing agents. The coatings were mainly of calcium carbonate, whereas phosphate is the mineral responsible for fossilization of the ancient eggs. But Martin *et al.* point out that, once the egg's outer



Spot the similarity: fresh (far left) and mineralized lobster eggs. Scale bars, 500 μm .

surface has become stable, preservation with phosphate may then occur over longer periods of time.

Tim Lincoln

Neurobiology

Stem cells stop dead

Neuron **37**, 209–219 (2003)

Some stem cells give rise to more cells in culture than they do during normal development. So how do they know when to stop dividing? Bruno C. Bello *et al.* find that, at least in parts of the developing fruitfly, precisely timed cell suicide is the answer.

In fruitfly larvae, neuronal precursors in the abdomen normally produce an average of nine new cells. But when Bello *et al.* prevented the precursors from committing suicide, 34 cells were generated instead. A pulse of Abdominal-A (AbdA) protein — one of the family of Hox developmental proteins, which distinguish head from tail — normally triggers the cell death. The team genetically engineered larvae to produce a premature burst of AbdA, which killed the precursors and slashed their progeny number.

The study is one of the first to show that Hox proteins directly affect cell division. Researchers now hope to find out whether related proteins that are active in the developing vertebrate nervous system might also control stem-cell proliferation by triggering well-timed death.

Helen Pearson

Fluid dynamics

Turbulence made hard

Phys. Rev. Lett. **90**, 034502 (2003)

The elusive state of convection known as 'ultrahard turbulence' is real, claim Detlef Lohse and Federico Toschi.

Convection — the rising of warm, buoyant fluid through cooler fluid — occurs in the atmosphere, the oceans, the deep Earth and many industrial processes. The fluid motion becomes turbulent when

'driven' hard, for example by sufficiently strong heating from below. Heat transport from bottom to top is related to the Rayleigh number, Ra . Theory predicts that the heat transport is proportional to $Ra^{2/7}$; experimentally, exponents of between 1/4 and 1/3 are found. At very high Rayleigh number, however, it has been predicted that the scaling should change to $Ra^{1/2}$ — which matters if, for example, you are trying to predict heat extraction in a nuclear power plant.

Lohse and Toschi have simulated convection in a way that eliminates effects at the boundaries of the fluid, and see the $Ra^{1/2}$ law at rather modest values of Ra . They conclude that an experiment might reach this regime if the influence of boundary layers could be eliminated.

Philip Ball

Evolution

Aquatic life and the larger male

Evol. Ecol. Res. **5**, 105–117 (2003)

In nearly all species of spider, females are much larger than males. One possible explanation — that males are miniaturized for mobility — now receives support from studies of *Argyroneta aquatica*, in which males are 30% bigger than females. This is the only spider to live entirely in water; the animals weave an underwater web and fill it with air, like a diving bell, returning to the surface to replenish it. Dolores Schütz and Michael Taborsky find that male *Argyroneta* are more active and better divers than females. This, they believe, is partly a consequence of their size, as larger animals overcome water resistance more easily. Males have longer front legs than females and a streamlined shape, which may also be adaptations for diving.

When it comes to terrestrial mobility, smaller size may be preferable because less work needs to be done against gravity. Males tend to travel to find mates, whereas females typically stay put. Other researchers have suggested that the evolutionary pay-off for larger females is increased fecundity.

John Whitfield

Patagonian toothfish found off Greenland

This catch is evidence of transequatorial migration by a cold-water Antarctic fish.

A large (1.80 metres in length and weighing 70 kg) Patagonian toothfish (*Dissostichus eleginoides* Smitt, 1898) has been caught in the northwest Atlantic, representing the first Northern Hemisphere record of the diverse, abundant and mainly Antarctic suborder Notothenioidei. This extraordinary catch indicates that large, cold-temperate fishes may occasionally migrate from sub-Antarctic to sub-Arctic waters by using deep, cold water, supporting a widely accepted but unproven proposal that the anti-tropical distribution patterns of many marine biota could be explained by transequatorial migration^{1–3}.

During commercial gill-net fishing for Greenland halibut in the Davis Strait (63° 02' N, 53° 32' W, at a depth of 1,331 m), a large, unknown fish was caught on 23 November 2000. Olaf Sólsker, the captain of the vessel *Isak L*, decided to freeze this strange specimen. We examined the fish and identified it as a male Patagonian toothfish, which is well known for being controversially caught by trawl and long-line fisheries in sub-Antarctic waters⁴.

None of the 139 notothenioid species has previously been recorded in the Northern Hemisphere⁵, but morphological analysis of the specimen (for example, IX and 28 rays in the dorsal fins, 28 in the anal fin and 26 in the pectoral fins) is consistent with data⁶ for *D. eleginoides*, except in a few details: the mid-lateral line does not extend forwards to the tip of the pectoral fin (Fig. 1), although the number of middle lateral-line scales (74) is within the range for *D. eleginoides* (64–77) and outside the range (35–48) for its sister species *D. mawsoni* Norman, 1937; there are two scaleless areas on the head as opposed to 12; the lower jaw has a short inner row of small teeth and a long outer row of large teeth, rather than a single row; in the upper jaw there are four rows and not two; the pectoral fin is shorter (12% of standard length, rather than 17–26%); and the body is deeper (23% of standard length, rather than 16–21%). These minor differences can be explained by allometric growth and are not indicative of a separate population or species. The only stomach content was an unidentifiable cephalopod upper beak.

It is unlikely that an unknown population of toothfish exists off Greenland, as none has been recorded during intensive fishing and surveying activity over the past 15–20 years. The species is already known to migrate over long distances (up to 1,800 km has been recorded by tracking⁷), and we consider it more likely that this was a stray specimen from a south Atlantic population.

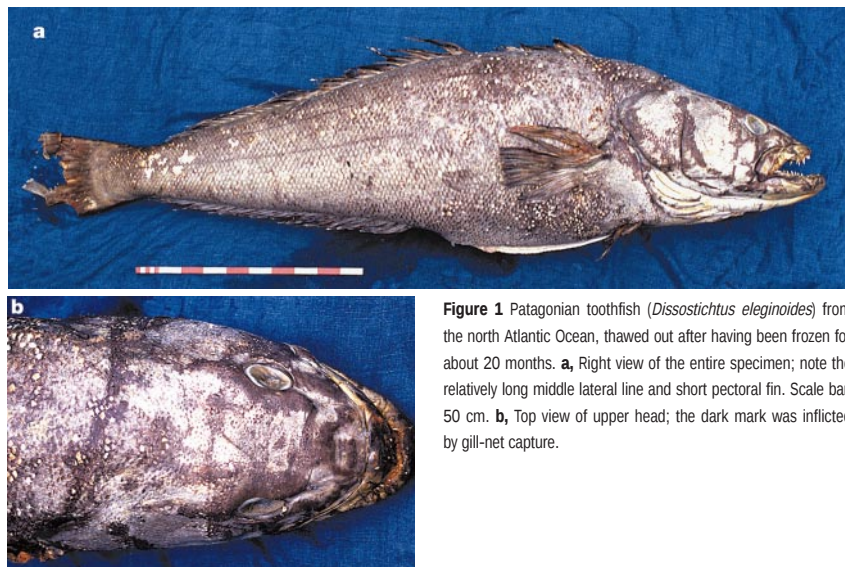


Figure 1 Patagonian toothfish (*Dissostichus eleginoides*) from the north Atlantic Ocean, thawed out after having been frozen for about 20 months. **a**, Right view of the entire specimen; note the relatively long middle lateral line and short pectoral fin. Scale bar, 50 cm. **b**, Top view of upper head; the dark mark was inflicted by gill-net capture.

The furthest north that a Patagonian toothfish has been recovered before⁸ is in the Atlantic Ocean off Uruguay at about 32° S, indicating that that specimen must have migrated at least 10,000 km.

The temperature range of *D. eleginoides* is 2–11 °C (ref. 9), so the transequatorial migration of the Greenland specimen must have included submergence beneath the intervening warm tropical waters, where temperatures are less than 10 °C at depths of 500–1,500 m (ref. 6); this strategy would enable the fish to travel all the way from Patagonia to west Greenland¹⁰.

Identified species of Greenland fish fauna have increased from 116 to more than 250 in 20 years, mainly as a result of increased fishing in deep waters. However, the discovery there of a Patagonian toothfish is so far unique. This finding supports the theory that transequatorial migration can occur by isothermal submergence^{1–3}, which has been proposed to explain the origin of marine anti-tropical distribution patterns.

Materials

Ultrahard polycrystalline diamond from graphite

Polycrystalline diamonds are harder and tougher than single-crystal diamonds and are therefore valuable for cutting and polishing other hard materials, but naturally occurring polycrystalline diamond is unusual and its production is slow. Here we describe the rapid synthesis of pure sintered

Peter Rask Møller*, **Jørgen G. Nielsen***, **Inge Fossen†**

*Zoological Museum, University of Copenhagen, 2100 Copenhagen Ø, Denmark
e-mail: pdr.moller@zmuc.ku.dk

†Møre Research, Section of Fisheries, PO Box 5075, 6021 Ålesund, Norway

- Ekman, S. *Zoogeography of the Sea* (Sidgwick & Jackson, London, 1953).
- Andriashev, A. P. *Polar Biol.* **11**, 213–218 (1990).
- Hilbish, T. J. et al. *Mar. Biol.* **136**, 69–77 (2000).
- Wienecke, B. & Robertson, G. *Fisheries Res.* **54**, 253–265 (2002).
- Balushkin, A. V. *J. Ichthyol.* **40** (suppl. 1), 74–109 (2000).
- DeWitt, H. H., Heemstra, P. C. & Gon, O. in *Fishes of the Southern Ocean* (eds Gon, O. & Heemstra, P. C.) 279–331 (J. L. B. Smith Inst. Ichthyol., Grahamstown, S. Africa, 1990).
- Williams, R., Tuck, G. N., Constable, A. & Lamb, T. J. *Sci. Committee Commission Conserv. Antarctic Mar. Living Res.* **9**, 33–48 (2002).
- Agnew, D. in *IUCN/SSC Wildlife Trade Programme* (Cambridge, UK, 2002).
- Eastman, J. T. in *Fishes of the Southern Ocean* (eds Gon, O. & Heemstra, P. C.) 34–51 (J. L. B. Smith Inst. Ichthyol., Grahamstown, S. Africa, 1990).
- Dietrich, G. & Ulrich, J. *Atlas zur Oceanographie* (Bibliographisches Inst., Mannheim, 1968).

Competing financial interests: declared none.

polycrystalline diamond by direct conversion of graphite under static high pressure and temperature. Surprisingly, this synthesized diamond is ultrahard and so could be useful in the manufacture of scientific and industrial tools.

We used a multi-anvil apparatus for high-pressure synthesis¹. A pure polycrystalline graphite (99.9995%) rod was found to transform into a transparent, colourless phase in only a few minutes at temperatures of 2,300–2,500 °C and at pressures of

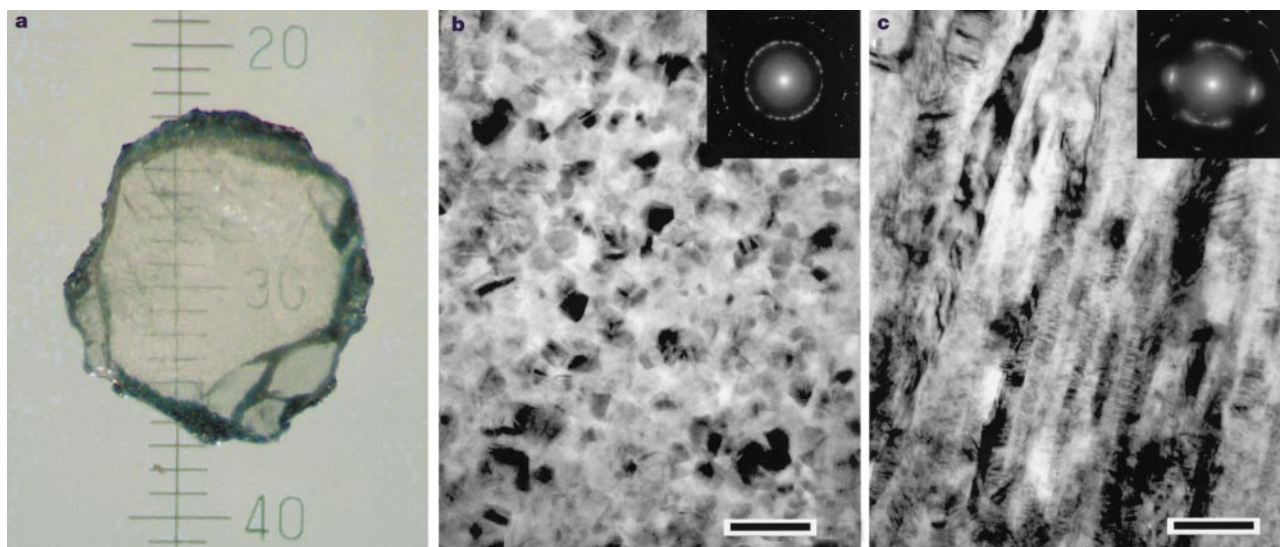


Figure 1 Sintered polycrystalline diamond synthesized by direct conversion of graphite. **a**, Optical microscopic image of a sample of the synthesized diamond (about 0.1 mm in diameter and 0.3 mm thick). **b**, **c**, Transmission electron microscopy reveals that this diamond material consists of minute crystals that are 10–20 nm across (**b**) and of larger, elongated (100–200 nm) crystals, which are evident in another region of the same sample (**c**). Scale bars, 50 nm. Insets, electron-diffraction patterns of each crystal type, obtained using a beam size of 400 nm.

12–25 gigapascals (GPa). X-ray diffraction and Raman spectroscopic measurements showed that this phase was pure cubic diamond. At lower temperatures (1,600–2,200 °C), a mixture of cubic and hexagonal diamonds, with a small amount of compressed graphite², was formed, which was opaque and dark grey in colour.

We studied a thin section taken from a transparent diamond sample (Fig. 1a) using a transmission electron microscope. The sample was seen to consist of very fine granular crystals, which were typically 10–20 nm across; electron diffraction indicated that these were randomly orientated (Fig. 1b). In other parts of the same sample, we observed elongated crystals of up to 100–200 nm in length that had a lamellar texture (Fig. 1c).

We measured the hardness of one opaque and two transparent samples, using a pure synthetic single-crystal diamond as a reference³. The transparent samples were found to have Knoop hardnesses of 110–130 GPa and 120–140 GPa, respectively, at several arbitrarily selected points, whereas the opaque sample was less hard (70–95 GPa).

Although diamond is the hardest known material, the Knoop hardness of single-crystal diamond varies from about 60–120 GPa, depending on the crystallographic plane and direction of measurement⁴. Our polycrystalline diamond is therefore as hard as (or even harder than) single-crystal diamond; moreover, this hardness is constant throughout the sample, unlike single-crystal diamond.

Previous attempts have been made to synthesize sintered polycrystalline diamond from graphite either by shock compression or by using static high pressure, but these were unsuccessful, mainly because the reaction time was too short and/or sample

heating was not homogeneous^{5–8}. Sintering diamond powders to produce a single large diamond also failed, because of heterogeneous stress distribution within the sample as a result of the extreme hardness of the raw diamond⁹.

Pure polycrystalline diamond has been produced previously by chemical-vapour deposition, but in a thin film and over several months to obtain millimetre-sized samples (see, for example, ref. 10). These diamonds were not sintered and had poor intergrain adhesion; the crystals were much larger than ours and their orientation was higher. They were accordingly less tough, with a hardness of 80–100 GPa (ref. 10).

Hard, sintered polycrystalline diamond is produced commercially by using a binder of metal or inorganic material^{11,12}, but this method reduces the hardness of the product to about 50–70 GPa. These sintered diamonds can be used only at temperatures of up to 600–700 °C — particularly in the presence of metal binders, which help diamond to transform back to graphite at ambient pressure. As our polycrystalline diamond is stable up to about 1,200 °C in an inert atmosphere, its hardness should exceed that of binder-containing polycrystalline diamond at high temperatures.

Improving the synthesis of the ultra-hard polycrystalline diamond described here could give rise to better-quality products and to new industrial applications, for example in scientific instruments that operate at high pressure. A better understanding of the process by which our diamond is formed should also provide insight into the enigmatic origin of natural polycrystalline diamonds¹³.

Tetsuo Irifune*, **Ayako Kurio***, **Shizue Sakamoto***, **Toru Inoue***, **Hitoshi Sumiya†**

**Geodynamics Research Center, Ehime University, Matsuyama 790-8577, Japan*

e-mail: irifune@dpc.ehime-u.ac.jp

†Itami Research Laboratories, Sumitomo Electric Industries, Itami 664-0016, Japan

1. Kawai, N. & Endo, S. *Rev. Sci. Instr.* **41**, 1178–1181 (1970).
2. Bundy F. P., Hall, H. T., Strong, H. M. & Wentorf, R. H. *Nature* **176**, 51–55 (1955).
3. Sumiya, H., Toda, N. & Sato, S. *Diamond Rel. Mater.* **6**, 1841–1846 (1997).
4. Brookes, C. A. & Brookes, E. J. *Diamond Rel. Mater.* **1**, 13–17 (1991).
5. Bundy, F. P. *J. Chem. Phys.* **38**, 631–643 (1963).
6. Naka, S., Horii, K., Takeda, Y. & Hanawa, T. *Nature* **259**, 38–39 (1976).
7. DeCarli, P. S. *Mater. Res. Soc. Symp. Proc.* **383**, 21–31 (1995).
8. Yusa, H. *et al. Appl. Phys. Lett.* **72**, 1843–1845 (1998).
9. Hall, H. T. *Science* **169**, 868–869 (1970).
10. Krauss, A. R. *et al. Diamond Rel. Mater.* **10**, 1952–1961 (2001).
11. Katzman, H. & Libby, W. F. *Science* **172**, 1132–1133 (1971).
12. Akaishi, M. & Yamaoka, S. *Mater. Sci. Eng.* **A209**, 54–59 (1996).
13. De, S., Heaney, P. J., Vicenzi, E. P. & Wang, J. *Earth Planet. Sci. Lett.* **185**, 315–330 (2001).

Competing financial interests: declared none.

correction

Memory enhancement in early childhood

C. Liston, J. Kagan

Nature **419**, 896 (2002)

In emphasizing references pertaining to brain maturation rather than to behavioural development, our brief list implied that this study was the first to demonstrate the emergence of long-term memory for events in infants. On the contrary, a large body of work published by P. Bauer, among many others, addresses this issue exactly and also forms the basis of the methodology developed for our study. Our contribution was to assess retention in children at three different ages after a four-month delay in order to test the *a priori* prediction, based on earlier studies of brain maturation, that nine-month-olds would be compromised relative to children in their second year. More extensive citation of work in this area was not possible because of the limited number of references permitted.

The DNA sequence and analysis of human chromosome 14

Roland Heilig^{*†}, Ralph Eckenberg^{*†}, Jean-Louis Petit^{*†}, Núria Fonknechten^{*†}, Corinne Da Silva^{*†}, Laurence Cattolico^{*}, Michaël Levy^{*}, Valérie Barbe^{*}, Véronique de Berardinis^{*}, Abel Ureta-Vidal^{*}, Eric Pelletier^{*†}, Virginie Vico^{*}, Véronique Anthouard^{*}, Lee Rowen[‡], Anup Madan[‡], Shizhen Qin[‡], Hui Sun[§], Hui Du[§], Kymberlie Pepin[§], François Artiguenave^{*}, Catherine Robert^{*}, Corinne Cruaud^{*}, Thomas Brüls^{*}, Olivier Jaillon^{*†}, Lucie Friedlander^{*}, Gaëlle Samson^{*†}, Philippe Brottier^{*}, Susan Cure^{*}, Béatrice Séguens^{*}, Franck Anière^{*}, Sylvie Samain^{*}, Hervé Crespeau^{*}, Nissa Abbasi[‡], Nathalie Aiach^{*}, Didier Boscus^{*}, Rachel Dickhoff[‡], Monica Dors[‡], Ivan Dubois^{*}, Cynthia Friedman[‡], Michel Gouyvenoux^{*}, Rose James[‡], Anuradha Madan[‡], Barbara Mairey-Estrada^{*}, Sophie Mangenot^{*}, Nathalie Martins^{*}, Manuela Ménard^{*}, Sophie Oztas^{*}, Amber Ratcliffe[‡], Tristan Shaffer[‡], Barbara Trask[‡], Benoit Vacherie^{*}, Chadia Bellemere^{*}, Caroline Belser^{*}, Marielle Besnard-Gonnet^{*}, Delphine Bartol-Mavel^{*}, Magali Boutard^{*}, Stéphanie Briez-Silla^{*}, Stéphane Combette^{*}, Virginie Dufossé-Laurent^{*}, Carolyne Ferron^{*}, Christophe Lechaplais^{*}, Claudine Louesse^{*}, Delphine Muselet^{*}, Ghislaine Magdelenat^{*}, Emilie Pateau^{*}, Emmanuelle Petit^{*}, Peggy Sirvain-Trukniewicz^{*}, Arnaud Trybou^{*}, Nathalie Vega-Czarny^{*}, Elodie Bataille^{*}, Elodie Bluet^{*}, Isabelle Bordelais^{*}, Maria Dubois^{*}, Corinne Dumont^{*}, Thomas Guérin^{*}, Sébastien Haffray^{*}, Rachid Hammadi^{*}, Jacqueline Muanga^{*}, Virginie Pellouin^{*}, Dominique Robert^{*}, Edith Wunderle^{*}, Gilbert Gauguier^{*}, Alice Roy^{*}, Laurent Sainte-Marthe^{*}, Jean Verdier^{*}, Claude Verdier-Discala^{*}, LaDeana Hillier[§], Lucinda Fulton[§], John McPherson[§], Fumihiko Matsuda^{||}, Richard Wilson[§], Claude Scarpelli^{*}, Gábor Gyapay^{*}, Patrick Wincker^{*}, William Saurin^{*}, Francis Quétiér^{*†}, Robert Waterston[§], Leroy Hood[‡] & Jean Weissenbach^{*†}

^{*} Genoscope-Centre National de Séquençage, [†] UMR-8030, CNRS et Université d'Evry, || Centre National de Génotypage, 91000, Evry, France

[‡] Institute for Systems Biology, Seattle, Washington 98103, USA

[§] Genome Sequencing Center, Washington University School of Medicine, St Louis, Missouri 63108, USA

Chromosome 14 is one of five acrocentric chromosomes in the human genome. These chromosomes are characterized by a heterochromatic short arm that contains essentially ribosomal RNA genes, and a euchromatic long arm in which most, if not all, of the protein-coding genes are located. The finished sequence of human chromosome 14 comprises 87,410,661 base pairs, representing 100% of its euchromatic portion, in a single continuous segment covering the entire long arm with no gaps. Two loci of crucial importance for the immune system, as well as more than 60 disease genes, have been localized so far on chromosome 14. We identified 1,050 genes and gene fragments, and 393 pseudogenes. On the basis of comparisons with other vertebrate genomes, we estimate that more than 96% of the chromosome 14 genes have been annotated. From an analysis of the CpG island occurrences, we estimate that 70% of these annotated genes are complete at their 5' end.

The draft sequences of the human genome^{1,2} have provided an unprecedented wealth of information on our genome, and have facilitated the identification of genes involved in human diseases. These drafts, however, contain a number of inconsistencies and gaps that have to be resolved to obtain a reliable molecular infrastructure on which we can anchor the entire set of human genes including their transcriptional start and stop signals, exons, splicing variants and regulatory elements, as well as the sequence variations found in various human populations. Nearly complete sequences of chromosomes 22 (ref. 3), 21 (ref. 4) and 20 (ref. 5) have already been achieved in the last three years. As an additional contribution to this goal, we present here the sequence of the euchromatic region of human chromosome 14. This chromosome contains two one-megabase (Mb)-long regions of prime importance for the immune system—the α/δ T-cell receptor (TCR) locus located close to the centromere, and the immunoglobulin heavy chain (IGH) locus adjacent to the telomere—as well as about 60 genes which, when defective, are known to lead to genetic diseases, including spastic paraplegia, Niemann–Pick disease, early onset Alzheimer's disease and a severe form of Usher syndrome.

The sequence quality reaches the internationally adopted standard of 99.99%. We estimate that we cover more than 99.9% of the euchromatic portion of chromosome 14, which, as in the other acrocentric chromosomes, namely 13, 15, 21 and 22, is restricted to its long arm. Particular care has been devoted to evaluating the integrity and accuracy of clone coverage. For its annotation, this sequence has benefited from the availability of the continuously

increasing amount of both genomic and expressed sequence data from humans and other vertebrates.

Sequence assembly and validation

Construction of the tiling path of DNA fragments (clones and polymerase chain reaction (PCR) products) that were used as sequencing material has been described previously⁶. Resolution of two gaps that remained in this sequence-ready map and of a few additional problems that appeared during the validation procedure are detailed in the Supplementary Information.

The final sequence-ready map consisting of 624 bacterial artificial chromosome (BAC) clones, 33 P1-derived artificial chromosome (PAC) clones and one P1 clone, segments of 8 yeast artificial chromosomes (YACs; subcloned into miniBACs or cosmids), 9 individual cosmid and 6 genomic PCR products spanned the long arm of chromosome 14 with no apparent gaps. A contiguous sequence assembly of this sequence-ready map was established and is presented in Fig. 1. On its centromeric end, the sequence contig begins in a highly conserved segmental duplication. This segmental duplication is highly homologous to a copy located in a similar position at the proximal end of the sequence of chromosome 22q^{3,7}. The distal end terminates in a unique sequence, which was mapped 5–8 kilobases (kb) from the telomere on the basis of half-YAC yRM2006 (<http://www.wistar.upenn.edu/Riethman/>) and exonuclease Bal-31 digestion⁸. We therefore conclude that the euchromatic part of chromosome 14q has been entirely sequenced. The mean clone overlap in the tiling path is 29.9 kb (22.5% clone redundancy)

and the mean finished sequence overlap is 4.7 kb (3.6% sequence redundancy).

To identify inconsistencies, we checked continuity, ordering and integrity of the clone and sequence coverage along the whole sequence assembly at several levels of resolution. This validation process is based on two rationales. First, the presence in the sequence of a number of reference points, provided by the marker collections from various maps and, conversely, the absence of sequence-tagged sites (STSs) belonging to other chromosomes. Second, the correlation of the restriction map derived from the sequence assembly with the restriction fragment pattern observed in non-sequenced BACs covering almost the entire chromosome 14 sequence (see Methods).

We first checked for integration of all the markers from two genetic maps and from the HAPPY map⁹ (see below and Supplementary Information). We also compared the marker content of the sequence to a radiation hybrid map⁶ comprising a denser set of 2,350 markers built on the TNG whole-genome radiation hybrid panel¹⁰ and including most markers from this radiation hybrid map. In all instances where markers were apparently missing from the sequence assembly, their location on chromosome 14 could not be confirmed by *de novo* mapping on radiation hybrids or single chromosome hybrids, and most of such markers could be positioned elsewhere in the genome.

Collinearity between the fingerprint contig (FPC)-based physical map¹¹ and the sequence map was also verified (see Supplementary Methods). Finally, we developed a procedure to validate the local clone and sequence assemblies as well as the integrity of the clones selected for sequencing. In this procedure we compared the restriction maps (for four enzymes) deduced from the assembled sequence to the experimental restriction patterns derived from an alternative set of BAC clones that was independent of the sequence-ready map but covered the sequence (see Methods). Ninety-three per cent of the sequence assembly, corresponding to the fraction covered by the alternative set of BACs, could be validated using this procedure (Fig. 1). For regions weakly represented in the two major BAC libraries used and for which we were unable to identify overlapping clones (representing another 5%, see Fig. 1), the validation was limited to the sequenced BAC itself. We ascribed the remaining 2% to experimental failures, escape of some regions from the resolution limits of the restriction pattern analysis, the presence of bacterial transposons in some alternative BACs, or the occurrence of indels (insertion/deletion) or restriction site polymorphisms.

Care was taken to ensure the integrity of the sequence assembly by independent means. However, none of the validation processes used is absolute, particularly in resolving very recent duplications extending beyond the BAC unit, and we cannot formally exclude that a very peculiar genomic situation could escape our validation procedure.

Matching the sequence to chromosome maps

We compared positions of the Génethon microsatellite markers¹² between the sequence of chromosome 14 and the genetic linkage maps from Génethon and deCODE¹³ (Fig. 2). All the Génethon markers could be found on the final sequence assembly. Notably, a region showing an unusually high recombination rate (recombination 'jungle') in female meiosis occurs in a region of 0.7 Mb in the middle third of the genetic maps. The genetic distances of the female meiosis map appear more compact for the deCODE map, especially in the recombination jungle region.

As chromosome 14 is the only human chromosome that was mapped by the HAPPY mapping procedure, it was of particular interest to compare the HAPPY map⁹ to the sequence map (Fig. 3b). In the HAPPY mapping technique, large genomic DNA segments, generated by random breakage, are separated by limiting dilution and analysed for their marker content. Physically close markers tend to segregate together on the same segments. The frequency of

co-segregation provides a measure of distance and information on marker order that can be used to establish maps¹⁴.

The HAPPY map was established using a large set of markers (1,001), of which ten have been reassigned to other chromosomes (dots on ordinate of Fig. 3b). Except for ten other cases that appear as isolated dots, the distances between markers with conflicting order are below 1 Mb, and most frequently below 100 kb. Of note, 63% of the markers were positioned on the HAPPY map in the same order as they occur on the sequence. Such a coincidence obtained with a map showing a mean marker spacing of approximately 85 kb is outstanding and much better than that observed for the radiation hybrid maps. In many instances, orders at the 10-kb range were in agreement, demonstrating the high-resolution power of HAPPY mapping. A quasi-linear correlation between the HAPPY map and sequence distances can be seen for most of the chromosome (Fig. 3b). A more detailed analysis is provided in Supplementary Information.

The HAPPY markers of chromosome 14 were distributed between sequence positions 2,555,234 and 88,773,851, and ranged from no marker in a few intervals larger than 1 Mb, such as between positions 65,037,628 and 67,071,623, to a density of 47 markers per Mb between positions 55.1 and 56.1 Mb. As expected, the distribution was clearly non-uniform, with some marker gaps in the 1-Mb range scattered over the chromosome and which largely reflected the distribution of Alu repeats along the chromosome (Fig. 3a). A more detailed analysis is provided in Supplementary Information.

Sequence variations

Sequence overlaps between consecutive BACs have been analysed for sequence variations, focusing on the large insertion/deletion events (indels) affecting segments greater than 100 base pairs (bp) such as insertions of Alu elements that are stably inherited and provide powerful landmarks for population and human evolution studies. The cumulative size of overlapping segments analysed amounts to approximately 19.7 Mb, of which 12.8 Mb originated from different haplotypes, in which a total of 14 large indels were

Figure 1 Overview of the euchromatic portion of human chromosome 14 with the centromere on the left and the telomere on the right. The following sections appear in the order from top to bottom: (1) a cytogenetic map positioning the cytogenetic bands on the sequence map; (2) positions of the Génethon microsatellite markers; (3) tiling path of the sequenced genomic clones designated by their GenBank/EMBL/DBJ accession numbers (the length of the double bar beneath is proportional to the length of the clone sequences that were used for the construction of the final finished sequence). Bacterial artificial chromosomes (BACs) from the RPCI-11 library are in black; BACs from the CITD library are in green; genomic polymerase chain reaction (PCR) products are in red; and clones from other origins (see text) are in blue. (4) Experimental sequence validation by the program WATSON (see text) using non-sequenced BACs (green areas) or sequenced clones (yellow areas). Segmental duplications on chromosome 14 with (5) intrachromosomal segmental duplications (blue) and (6) interchromosomal segmental duplications (red)—the thinnest bars represent sequence matches between 1,000 and 8,000 bp. (7) The (G + C) content is expressed in per cent and is represented by a red curve along with a horizontal line representing the average (G + C) percentage (41%) of chromosome 14. The densities of long interspersed nucleotide elements (LINE) and short interspersed nucleotide elements (SINE) expressed by their cumulative length in an interval of 50 kb (%) are represented by an aquamarine and a black line, respectively. Exons are highlighted by filled green vertical bars, the heights of which correspond to cumulative exon length in an interval of 50 kb (%). The scale is in megabases (Mb); '0' corresponds to a position 1,550,000 bp upstream of the sequenced euchromatic start position (see Supplementary Methods). (8) The bottom part of the figure highlights the gene content of the sequence. The annotated gene categories, as described in the text, are represented by coloured bars. The width of the bars is proportional to the gene length. A horizontal strip separates upper- and lower-strand genes, and includes positions of the CpG islands, indicated as vertical bars. (9) The TCR and IGH loci are indicated by boxes in the last two lines.

detected. Among these, eight were Alu insertion polymorphisms (see Supplementary Information) observed in one of two sequenced haplotypes, indicating that Alu insertions account for the most abundant source of large insertion polymorphisms in the human genome. Assuming that chromosome 14 is representative of the human genome, an extrapolation of these events to the entire genome provides a first rough estimate of Alu insertion polymorphisms in the human genome of about 0.7 insertions per million base pairs when matching two haplotypes. Other large indels include a 60-kb segment and are described in Supplementary Information.

The chromosome 14 landscape

The chromosome 14 sequence shows a (G + C) content of 40.86%, which agrees almost exactly with the value (41%) reported for the entire genome. Genes cover a total of 38.069 Mb (43.6% of the total sequence), for a cumulated exon fraction of 2.021 Mb (2.3%) and a potential coding fraction of 1.000 Mb (1.1%). Total repetitive elements cover 40.373 Mb (46.2%), distributed essentially in short interspersed nucleotide elements (SINEs) (11.620 Mb) and large interspersed nucleotide elements (LINEs) (17.327 Mb), which corresponds to 13.3% and 19.8% of the whole sequence, respectively. These characteristics are compared with other finished chromosomes in Table 1. However, the chromosome 14 landscape is far from uniform (see Fig. 1): important fluctuations in (G + C) content occur, between 32.6% and 61.2% (for a window size of 50 kb), with high (G + C) content narrowly correlated with a dense CpG island distribution, a high SINE (versus low LINE) content and a high exon density. At a global level, variations in the (G + C) content occur in a discrete pattern, juxtaposing regions of very different base composition, such as isochore-like domains¹⁵. The gene distribution along the chromosome follows the same discontinuity, with a succession of gene-rich islands tightly superimposed over peaks of SINE elements (and also LINE depressions) and separated by very gene-poor areas; the five longest of them cover 2–6 Mb. Gene organization in gene-dense regions seems to occupy the chromosomal space available in an optimal manner, which is exemplified by a quasi-continuous succession of condensed gene structures often arranged in a forward and reverse alternance so that CpG islands in the region are often shared by genes of inverted polarity. These regions are often followed, with little or

no transition, by long stretches depleted in exons or occupied by very large genes with big introns. For example, the 2.7-Mb-long region between positions 55.2 Mb and 57.9 Mb, which shows a high exon density (5.6%, that is 20.7 genes per Mb), has a global (G + C) composition of 45.2%, with several peaks over 50%, and presents a SINE density of 29.3% versus a LINE density of 10.7%, values that are quite deviant from their respective averages.

A total of 1,768 CpG islands were found along the chromosome sequence, more than 530 of which appear to be related to the 5' end of genes. They are especially concentrated in the subcentromeric and subtelomeric regions (up to one CpG island per 9 kb, compared with the chromosome average of one per 50 kb). These two regions contain an olfactory receptor gene cluster, the α/δ TCR locus (subcentromeric region) and the IGH locus (subtelomeric region), and are rich in potential coding segments and in non-processed pseudogenes. On the contrary, the region between 65.1 Mb and 66.2 Mb, which corresponds to the biggest 'gene desert' (no annotated exon in 1.111 Mb, compared with 43.5 kb for the average intergenic region), shows a low (G + C) content (35.2%) and a weak SINE density (6.6%). This desert also seems to be depleted in CpG islands (one per 75 kb). Three gene deserts larger than 1 Mb and 15 larger than 500 kb cover almost 11.7 Mb (13.4%) of the chromosome 14 landscape.

An analysis of the synteny between human chromosome 14 and its mouse counterpart, at both genomic and gene levels, is reported in the Supplementary Information. The distribution (in position and orientation) of mouse genome segments (essentially from murine chromosomes 12 and 14) that are syntenic with human chromosome 14 are represented in Supplementary Fig. 2.

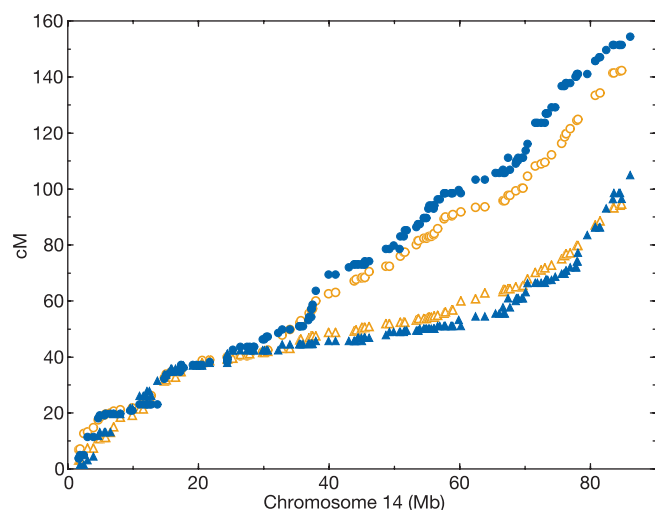


Figure 2 Comparison between genetic linkage maps and the chromosome 14 sequence. For genetic linkage maps, upper curves (circles) represent female recombination maps (filled circles, Génethon map; open circles, deCODE map), and lower curves (triangles) represent male recombination maps (filled triangles, Génethon map; open triangles, deCODE map).

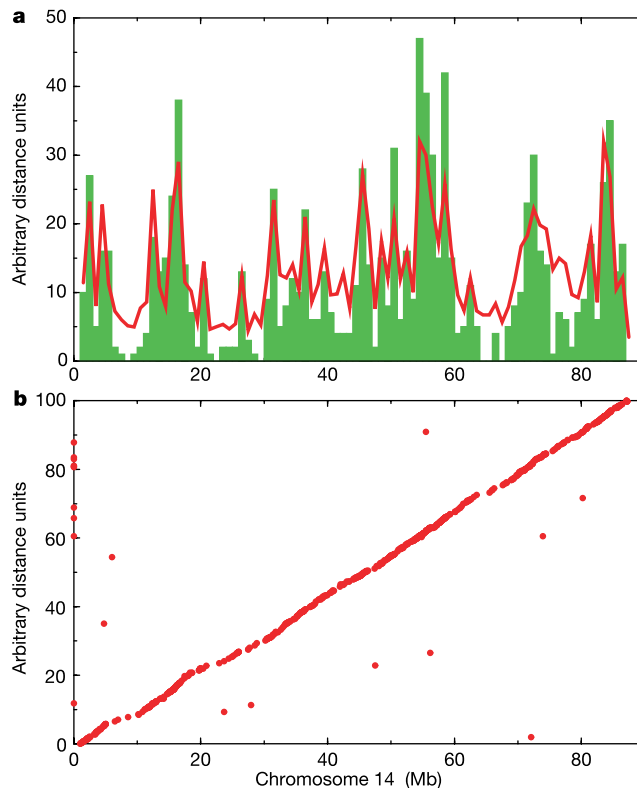


Figure 3 Comparison between the HAPPY map and the sequence of human chromosome 14. **a**, Distribution of the HAPPY map markers (histogram) and Alu elements (curve) per Mb along chromosome 14. **b**, Representation of the HAPPY map on the chromosome 14 sequence. Each dot represents the position of a marker on both the HAPPY map and the sequence assembly. Dots on the ordinate correspond to markers not found in the chromosome 14 sequence, but are found instead on other human chromosomes.

Table 1 Comparison of chromosome features

Genome property	14	20	Chromosome 21	Chromosome 22	Whole genome
(G + C) content (%)	40.9	44.1	40.9	47.8	41.0
Repetitive elements (%)	46.2	42.0	40.1	41.9	44.8
Gene coverage (%)	43.6	42.4	—	39.0	27.0
Exon coverage (%)	2.3	2.4	—	3.0	—
Gene density (genes per Mb)	10.0	12.2	6.7	16.3	9.3–10.8
Known genes mean size (kb)	58.7	51.3	57.0	—	27.0
All genes mean size (kb)	45.7	34.7	39.0	—	—
Pseudogenes (%)	26.0	18.8	20.8	19.7	—
Alternative splicing (%)	54.0	35	—	59	—
Putative genes (%)	25.0	16	(*)	(*)	—

Statistics on gene categories of sequenced chromosomes or whole-genome draft sequence¹. (*) Indicates that a different definition of putative genes was used for chromosomes 21 and 22.

Segmental duplications

An ancient duplication involving 70% of chromosome 14 and a portion of chromosome 2 has been reported². This event is, however, only visible at the protein level and predates the mouse–human separation. In our study we focused on more recent events that are detectable at the DNA level.

Using genome-wide sequence alignments (see Supplementary Methods), we found that 1.6% of chromosome 14 consists of interchromosomal segmental duplications contained in fragments of 1 kb or more that show at least 90% sequence identity. A comparable value, based on a different comparison procedure, was reported earlier¹⁶, and confirms that chromosome 14 has the lowest content of interchromosome segmental duplications in the human genome. These duplications are scattered along chromosome 14, with an increased density in the proximal third and some areas of high concentration in the subcentromeric and subtelomeric regions (Fig. 1), which contains members of the TCR and immunoglobulin superfamilies. The largest duplicated segment is shared with chromosome 22. As already shown¹⁷, these paralogous segments are adjacent to the centromeres of the two chromosomes and show a sequence identity above 98%.

Very divergent values for internal segmental duplications, ranging from 0.6%¹ to 4%¹⁶ have been reported previously. It should be noted, however, that an undetermined fraction of sequence redundancy characterized the early genome assemblies, which increases the predicted values of intrachromosomal segmental duplications. In a similar analysis that excluded repetitive DNA, we found that internal duplications account for 1.1% of chromosome 14 and were

clustered in four segments (Fig. 1). The largest includes an 800-kb region adjacent to the centromere, which is also part of the segmental duplication shared with chromosome 22.

RNA genes and mitochondrial DNA insertions

A total of 14 of the 20 transfer RNA genes identified on chromosome 14 using the tRNAscan-SE program are clustered in a segment of 75 kb (positions 2.947 to 3.021 Mb), which appears as the most dense cluster of transfer RNA genes in the human genome¹. Several small non-coding RNAs were localized along the chromosome sequence, including a complex tandem organization in an imprinted domain in 14q32. No complete ribosomal RNA-coding gene was found on the chromosome 14 long arm. Finally, 11 mitochondrial DNA insertions were characterized (see Supplementary Information for details).

Gene index of chromosome 14

Annotation of chromosome 14 results in 1,443 gene models that were classified into eight categories (see Table 2 and Methods): (1) 506 'known genes', identical to human complementary DNA or protein sequences identified by a LocusID in the LocusLink database (<http://www.ncbi.nlm.nih.gov/LocusLink>); (2) 110 'novel genes', having a coding sequence (CDS) and which are identical or homologous to spliced expressed sequence tags (ESTs) (vertebrates) or identical (but without a LocusID) or homologous to known cDNAs (vertebrates) and/or proteins (all species); (3) 11 'novel transcripts', similar to novel genes, but for which a unique CDS could not be defined unambiguously; (4) 212 'putative genes', homologous to cDNAs or spliced ESTs (vertebrates) but devoid of a significant CDS; (5) 11 'predicted genes', based on *ab initio* predictions and for which at least one exon is supported by biological data (unspliced ESTs, protein sequence similarities with mouse or *Tetraodon* genomes or expression data from Rosetta); (6) 296 'pseudogenes', sequences homologous to cDNAs or proteins (in at least 50% of the resource length) with a disrupted CDS and for which an active gene could generally be found at another locus. Finally, two specific categories were created for the annotation of the elements belonging to the TCR α/δ locus (positions 3.959 to 5.073 Mb) and the IGH locus (positions 87.685 to 88.954 Mb). These loci have an intrinsic germline organization correlated with inherent biological mechanisms essential for the generation of active genes (namely recombination events between ordered sets of segments), such that a classical exon/intron structure cannot be

Table 2 Distribution of genes among the different categories in human chromosome 14 annotations

Categories of genes	No. of genes	Gene length (nt)*	Transcript length (nt)*	No. of exons (exons per gene)	Exon length (nt)*	No. of alternative transcripts	CDS length (nt)*	CpG (–2,000 to 1,000 bp)
Known genes	506	58,748 (1,691,847)	3,080.3 (21,776)	5,206 (10.3)	299.4 (10,334)	287 (57%)	1,788 (20,658)	379 (75%)
Novel genes	110	35,818 (302,130)	1,901.3 (6,139)	765 (7.0)	273.4 (4,607)	51 (46%)	1,079 (5,616)	69 (63%)
Novel transcripts	11	13,316 (51,010)	2,326.4 (4,691)	32 (2.9)	799.7 (4,691)	1 (9%)	487 (954)	8 (73%)
Predicted genes	11	34,334 (170,900)	1,503.4 (4,227)	68 (6.2)	243.2 (2,770)	0	1,237 (2,217)	6 (55%)
Putative genes	212	22,005 (388,767)	732.0 (3,296)	600 (2.8)	258.6 (3,296)	23 (11%)	—	69 (33%)
All genes	850	45,713	2,311.9	6,671 (7.8)	294.6	362 (43%)	—	531 (62%)
Gene segments	200	581 (10,131)	280.4 (2,330)	—	—	6 (3%)	—	11 (5%)
Pseudogene segments	97	506 (5,462)	356.4 (2,870)	—	—	—	—	7 (7%)
Pseudogenes	296	1,431 (40,592)	1,244.5 (8,514)	—	—	—	—	—

Definitions of the categories are described in the text. CpG islands found in an interval spanning from 2,000 bp upstream to 1,000 bp downstream from the most 5' end of the annotated gene were scored. nt, nucleotides. CDS, coding sequence.

* Average gene-length values are shown, with longest gene-length values indicated in parentheses.

assigned for these genes. We therefore defined these structural elements as: (7) 200 'gene segments'; and (8) 97 'pseudogene segments' (when inactivated), and analysed these and the corresponding genomic regions independently. The gene index of human chromosome 14 is presented in Supplementary Table 4 and will be regularly updated at <http://www.genoscope.cns.fr/chr14/>. Gene names will be regularly reviewed by the Human Genome Organisation (HUGO) Gene Nomenclature Committee (HGNC) (<http://www.gene.ucl.ac.uk/nomenclature/>).

Excluding pseudogenes and 'segments', chromosome 14 presents a mean gene density of 10 genes per Mb (850 genes per 85 Mb), which is close to the mean value estimated for the whole human genome (9.3–10.8 genes per Mb¹). Table 1 gives a comparison of the gene density between the four finished chromosomes. Pseudogenes accounted for 26% of the total, which is slightly higher than for the other finished chromosomes (Table 1). The mean gene length is 45.7 kb on the basis of the first five categories above, with significant variations between them (see Table 2 and an interchromosomal comparison in Table 1). Six known genes exceed 500 kb, whereas 104 genes (82 known, 13 novel, 8 putative and one predicted) are over 100 kb in length. The largest gene, the neurexin 3 (*NRXN3*) gene¹⁸, extends over 1,691,847 bp and the largest messenger RNA is 21,776 bp long and consists of 115 exons that encode SYNE-2, a protein of 6,649 amino acids expressed in synaptic nuclei at the neuromuscular junction¹⁹.

Among the 506 known genes, three CDSs use internal TGA triplet(s) to encode selenocysteine, namely the glutathione peroxidase 2 (*GPX2*) gene²⁰ and the genes for deiodinase iodothyronine types II and III (*DIO2* and *DIO3*)^{21,22}. Seventeen known genes have no CDS longer than 100 amino acids. These genes might encode a short peptide product. Alternatively, as hypothesized for the maternally expressed genes *MEG3* and *MEG8* from the imprinted domain in 14q32 (ref. 23), which is syntenic to the ovine callipyge locus, some genes that have a clear mosaic structure but do not present a significant CDS might act by means of a non-coding mRNA-like element²⁴ rather than a protein product.

Notably, about a quarter of annotated genes (212 out of 850) were classified as putative, showing an exon/intron structure but lacking a significant CDS, according to our criteria. They correspond in part to incomplete transcription units, annotated from very partial resources and for which experimental determination of the complete structure will be needed. As discussed above, some of them might also correspond to genes encoding short peptides or act by means of a non-coding mRNA-like transcript.

We found 362 (43% of 850) annotated genes that have one or more alternative transcript. The prevalence of alternative splice events in each gene category is given in Table 2. Excluding the putative and predicted genes, which are usually incomplete, alternative splicing occurs in 339 genes (54% of 627). This value is comparable to that reported for chromosome 22 (59%)¹ and is noticeably higher than that reported for chromosome 20 (35%)⁵ (Table 1). Indeed, owing to our conservative approach in annotation of alternative transcripts, in which we did not discriminate for the presence or absence of a CDS, we cannot exclude the possibility that a fraction of rare splicing events might be due to splicing machinery background²⁵.

The number of distinct transcripts averages 2.5 per gene for the

'known' category; however, several genes exhibit extensive alternative splicing potential (see Supplementary Information). The most notable is the large neurexin 3 gene, that contains a functional internal alternative promoter. At five sites of the long form, which are reduced to two sites in the shorter form, large numbers of possible splicing variants were postulated, resulting in a theoretical set of 1,728 combinatorial possibilities for alternative transcripts¹⁸. Because the exact function of neurexin 3 is unknown, the reasons for the large gene size and extensive alternative splicing remain a mystery. Another interesting example is given by the imprinted genes *MEG8* and *MEG3*, for which five and eight alternative transcripts have been described respectively, although none of them seems to correspond to a CDS²³.

From an analysis of 10,164 splice junctions, we observed that minor types of splice sites (different from the canonical GT-AG pattern) account for less than 1.47% (see Supplementary Information for details), with a relative prevalence of GC-AG (87 cases) and AT-AC (5 cases).

The 5' end of annotated genes

To evaluate the fraction of complete 5' ends in annotated genes after human curation, we investigated the presence of CpG islands in the 5' regions of genes (−2 kb, +1 kb), for each category and determined its correlation with the annotation of unambiguous 'start' codons. Our *in silico* definition of an initiator methionine was conservative, on the basis of the presence of upstream in-phase 'stop' codon(s) (see Methods). Using this procedure, unambiguous start codons were annotated for 358 known genes (73%), of which 266 (74%) have their 5' end in the vicinity of a CpG island. Considering previous estimations that 60–67% of genes are associated with a CpG island at their 5' end^{5,26}, we assumed that most of the 92 known genes (26%) with an unambiguous start codon but not linked to a CpG island at their annotated 5' end are nevertheless complete, or nearly so. Similarly, 77% of the 130 known genes for which the program did not find an unambiguous initiator Met are associated with a CpG island in their annotated 5' end. We therefore considered that most (if not all) genes from the known category and, to a lesser extent from the novel categories are essentially complete at their annotated 5' end, because of the close proximity of a CpG island in 75% and 64% of the cases, respectively. Assuming correct identification of the 5' end of genes associated with CpG islands, an initiator methionine could be defined for 100 additional known and 33 novel genes. The 5' end of the putative and predicted genes however, is currently too hypothetical to be associated significantly with CpG islands. On the basis of these data, we estimate that about 90% of the annotated genes (excluding putative genes) or about 70% (including putative ones) should be complete.

Contribution of comparative genomics resources

The process of annotation includes comparisons derived from partially sequenced genomes of *Mus musculus* ('BLAT-mouse'; see Methods) and *Tetraodon nigroviridis* (Exofish²⁷). No gene could be identified on the basis of a comparison with the deduced proteome of *Caenorhabditis elegans* and *Drosophila melanogaster* alone. On the other hand, mouse and fish genome sequence comparisons contributed significantly to the refinement of gene annotation on

Table 3 Specificity and sensitivity of the non-expressed resources in the annotation of human chromosome 14

	Exofish	BLAT-mouse	GENSCAN exons	FGENESH exons	Exofish + BLAT-mouse	Exofish + BLAT-mouse + GENSCAN exons + FGENESH exons
Annotation features	4,396	18,583	10,192	7,901	3,862	2,609
Specificity*	96%	31%	52%	61%	96%	99%
Sensitivity†	42%	70%	68%	63%	74%	83%

* The fraction of the annotation features of the various resources or a combination of resources that were included in the exons of the final gene annotation is indicated.

† The total number of annotated exons for all gene categories is 7,305 exons. The fraction of exons identified with the various resources or with a combination of resources is indicated.

human chromosome 14 (5' or 3' extensions, fusions of gene structures, putative alternative exons) and provided support for annotation as predicted genes for 11 *ab initio* determinations (an example of the contribution of comparative genomics data is given in Supplementary Information). Although the precise structure of such determinations needs to be experimentally confirmed, this annotation already constitutes a working model.

We tried to evaluate the contribution of non-expressed resources to the detection and refining of gene structures retrospectively (see Supplementary Tables 1 and 2). Table 3 shows the values of specificity and sensitivity for Exofish, BLAT-mouse, GENSCAN²⁸ and FGENESH²⁹, in several combinations. Although Exofish might be insufficient to identify all of the exons despite its high specificity, particularly when they belong to rapidly evolving genes, its sensitivity attains 82% for the detection of the RefSeq transcripts and 84% for the known genes. This last sensitivity value reaches 96% when both genomic resources are combined (see Supplementary Table 2), providing an estimation of the completeness of the gene identification.

Missing genes

Models resulting from the gene prediction algorithms (GENSCAN and FGENESH) and supported, even partially, by non-expressed resources only (Exofish and/or BLAT-mouse) were examined and, in 11 occurrences, annotated as predicted genes. The remaining *ab initio*-only models include 116 (out of 586) GENSCAN-only, 70 (out of 157) FGENESH-only and 58 (out of 990) combined (that is overlapping in at least one exon) models. Although we cannot exclude the possibility that some of these could pin-point a missing gene (especially for the combined type), we estimate that they essentially represent a background that is intrinsic to the prediction algorithms. Indeed, the mean probability for a GENSCAN-predicted exon is 0.80 inside annotated exons versus 0.40 outside. Such *ab initio*-only models were excluded from the present annotation.

On the other hand, a fraction of single exons (evolutionarily conserved regions) (and, to a lesser extent owing to their lower specificity, of BLAT-mouse comparisons) that match outside annotated genes might correspond to exons of missing genes, particularly when conserved in all three species. To purge genomic DNA contamination of EST libraries we excluded unspliced ESTs devoid of CDS from annotation, as well. Such products could, however, correspond to some 5' or, most likely, 3' untranslated portions of actual transcripts. We are considering experimental testing of some of these products that are questionable individually, but for which the number of matching ESTs suggests biological relevance.

Finally, some genes that undergo rapid evolution, that are specific for minor cell types, or that show a very sharp pattern of expression or display an uncommon base composition might have escaped all current detection or prediction strategies.

Conclusion and medical implications

The chromosome 14 long-arm sequence was generated essentially by a sequence tag connector strategy (STC)⁶. This sequence covers, in a single 87,410,661-bp-long contig, the entire euchromatic portion of the chromosome, from the subcentromeric region to the telomere. We made a special effort to resolve cloning and sequencing gaps and to validate the whole assembly combining various mapping resources and experimental procedures. Comparisons to genomic or expressed sequence data from human and other organisms, combined with *in silico* gene predictions, allowed us to establish a gene index for human chromosome 14. This annotation, which represents the present state of knowledge, is available on our web site (<http://www.genoscope.cns.fr/chr14/>) and will be regularly updated, including results of experimental validation, especially those for the numerous putative genes. When extended to the whole genome, such a catalogue should be considered as an invaluable

informational basis for the resolution of the complex network of functions and interactions between genes and gene products.

Functional studies can also greatly benefit from mapping of biological activities and disease gene phenotypes that have accumulated during the past decades. We have therefore examined the OMIM entries localized to chromosome 14 for which no sequence data is yet available. This analysis (which is described in Supplementary Information) enabled us to clarify some mapping issues and to identify some alternative strategies for functional studies.

Two large gene clusters of immunological importance, namely the α/δ TCR and the immunoglobulin heavy chain, are located on chromosome 14. The complete structure and organization of these gene loci form the basis of our knowledge of the germline origins of TCR and immunoglobulin repertoires as well as the molecular pathology of lymphocyte malignancies that often involve chromosomal translocations within these gene loci. The availability of the complete sequence of chromosome 14 confirms that the α/δ TCR complex has already been exhaustively characterized. However, the sequence of the IGH constant region gene locus, now complete, will shed new light on the dissection of the molecular mechanisms of isotype switching and affinity maturation of immunoglobulin heavy-chain molecules.

Several disease gene identifications have benefited from the availability of draft or finished sequences from chromosome 14 BACs. These include autosomal dominant spastic paraplegia (SPG3A)³⁰, molybdenum cofactor deficiency³¹, oligodontia³², type I Leber congenital amaurosis³³, microphthalmia³⁴ and type C2 Niemann–Pick disease (NPC2)³⁵. Some 25–30 loci are not yet linked to sequences and represent future targets for positional cloning³⁶. Most correspond to mendelian disease phenotypes segregating in families. These diseases are quite diverse and include one locus for the very severe Usher syndrome (USH1A) and several other hearing and vision impairments. The identification of all these genes should greatly benefit from the complete sequence and thorough analyses that are now available. □

Methods

Sequence integrity assessment

The program WATSON (E. Pelletier, personal communication) constructs the chromosome sequence assembly from individual BAC sequences and deduces the restriction map from the master-sequence obtained; inventories all of the overlapping BACs from BAC-end sequence resources (see Supplementary Information) and proposes an alternative tiling path excluding clones selected for sequencing; reconstructs the clone sequence (including the specific vector in the correct orientation) and deduces a virtual restriction pattern for four selected enzymes; compares (in the 0.5–50 kb range) this virtual pattern to the experimental restriction pattern performed from BACs of this alternative tiling path; and scores the concordant results on a 0–4 scale (for non-redundant segments) or 0–8 (for redundant segments). Taking into account experimental biases, segments scoring 2 or more were considered as valid, in terms of assembly and of clone integrity. Discordant results were reanalysed after selection of other alternative overlapping BACs. This allowed resolution of experimental artefacts (owing to error in overlapping clone selection, non-integrity of overlapping BACs, namely partial deletions, bacterial insertion elements, and so on) and also pin-pointed the actual non-validated segments in the master sequence.

Annotation procedure for protein-coding genes

We developed a two-step annotation process. The first step consisted of *in silico* comparisons performed between the repeat-masked genomic sequence and expressed sequence resources (RNAs, ESTs and proteins) from human and other vertebrates, leading to the definition of preliminary gene models. Then, the structure of each gene model (including small exons and repeat-containing fractions) was refined by processing the SIM4 (ref. 37) alignment between the relevant transcribed sequences and the corresponding unmasked genomic segment (extended by 5 kb at both ends). In addition, comparative sequence analysis was performed against the genome (or proteome) of completely or partially sequenced organisms: *M. musculus* (BLAT³⁸ alignment coordinates extracted from the University of California, Santa Cruz genome annotation database and mapped on our chromosome 14 assembly); *T. nigroviridis* (Exofish²⁷); *D. melanogaster* (http://www.fruitfly.org/sequence/sequence_db/); and *C. elegans* (http://www.sanger.ac.uk/Projects/C_elegans/wormpep). Together with this homology analysis, *ab initio* gene predictions were performed using two algorithms: FGENESH²⁹ and GENSCAN²⁸. In addition, experimental data based on microarray expression measurements performed by Rosetta³⁹, which support GENSCAN predictions, were used. Finally, CpG islands were detected along the chromosome 14 sequence using GRAIL

(<http://compbio.ornl.gov/grailexp/>). In a second step, the preliminary gene structures defined automatically were submitted to manual curation, to incorporate additional data leading to the extension, fusion or splitting of gene models and to characterize alternative splicing events. Transcripts were summarized for each gene by their most complete representative (one for each form, in case of alternative splicing) and a 'proposed model' of that gene was constructed that included the additional data. Human expertise was also required to exclude from the final annotation suspicious data as unspliced ESTs and cDNAs or very partial matches, showing nonsignificant CDSs, to reduce the background owing to experimental artefacts (that is, genomic contaminants in cDNA libraries).

CDS determination

Exon sequences of a gene model were concatenated and start and stop triplets were identified. Each putative coding region extending from a start triplet (or the beginning of the first exon), to a stop triplet (or the end of the last exon) was selected if its length was at least 300 bp. The CDSs contained, in the same phase, in another CDS were eliminated and the proximal methionine was identified. If no stop codon was found in phase between this proximal methionine and the 5' extremity of the concatenated virtual cDNA, this extremity was considered as a provisional CDS starting point. In a last step, three kinds of scores were calculated on the remaining CDSs: (1) a score equal to the ratio of the length of individual CDSs versus the length of the longer one; (2) a score relative to the 5' position (the CDS starting at the 5' extremity of the gene model had a score equal to 1 and dropped to 0 when the CDS started at a 10,000-bp distance); (3) a score relative to coding exons (a CDS overlapping all the exons of a gene model received a score equal to 1). The presence of 1, 2, 3 and 4 non-coding exons in 5' ends (and of 1, 2 and 3 non-coding exons in 3' ends) decreased the score by 0.025, 0.075, 0.2 and 0.375 (and 0.075, 0.2 and 0.375) respectively. The mean of these three scores was the final score for a CDS. The CDSs with a final score greater than 0.6 were retained, and ultimately CDSs with final scores below 90% of the best CDS final score were rejected.

Received 20 August; accepted 3 December 2002; doi:10.1038/nature01348.

Published online 1 January 2003.

1. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
3. Dunham, I., Shimizu, N., Roe, B. A. & Chisoe, S. The DNA sequence of human chromosome 22. *Nature* **402**, 489–495 (1999).
4. Hattori, M. *et al.* The DNA sequence of human chromosome 21. *Nature* **405**, 311–319 (2000).
5. Deloukas, P. *et al.* The DNA sequence and comparative analysis of human chromosome 20. *Nature* **414**, 865–871 (2001).
6. Bröls, T. *et al.* A physical map of human chromosome 14. *Nature* **409**, 947–948 (2001).
7. Chen, C. & Birshstein, B. K. Virtually identical enhancers containing a segment of homology to murine 3' IgH-E(hs1,2) lie downstream of human IgC α 1 and C α 2 genes. *J. Immunol.* **159**, 1310–1318 (1997).
8. Wintle, R. F., Nygaard, T. G., Herbrick, J. A., Kvaloy, K. & Cox, D. W. Genetic polymorphism and recombination in the subtelomeric region of chromosome 14q. *Genomics* **40**, 409–414 (1997).
9. Dear, P. H., Bankier, A. T. & Piper, M. B. A high-resolution metric HAPPY map of human chromosome 14. *Genomics* **48**, 232–241 (1998).
10. Olivier, M. *et al.* A high-resolution radiation hybrid map of the human genome draft sequence. *Science* **291**, 1298–1302 (2001).
11. The International Human Mapping Consortium. A physical map of the human genome. *Nature* **409**, 934–941 (2001).
12. Dib, C. *et al.* A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**, 152–154 (1996).
13. Kong, A. *et al.* A high-resolution recombination map of the human genome. *Nature Genet.* **31**, 241–247 (2002).
14. Dear, P. H. & Cook, P. R. Happy mapping: linkage mapping using a physical analogue of meiosis. *Nucleic Acids Res.* **21**, 13–20 (1993).
15. Clay, O. & Bernardi, G. Compositional heterogeneity within and among isochores in mammalian genomes. II. Some general comments. *Gene* **276**, 25–31 (2001).
16. Bailey, J. A., Yavor, A. M., Massa, H. F., Trask, B. J. & Eichler, E. E. Segmental duplications: organization and impact within the current human genome project assembly. *Genome Res.* **11**, 1005–1017 (2001).

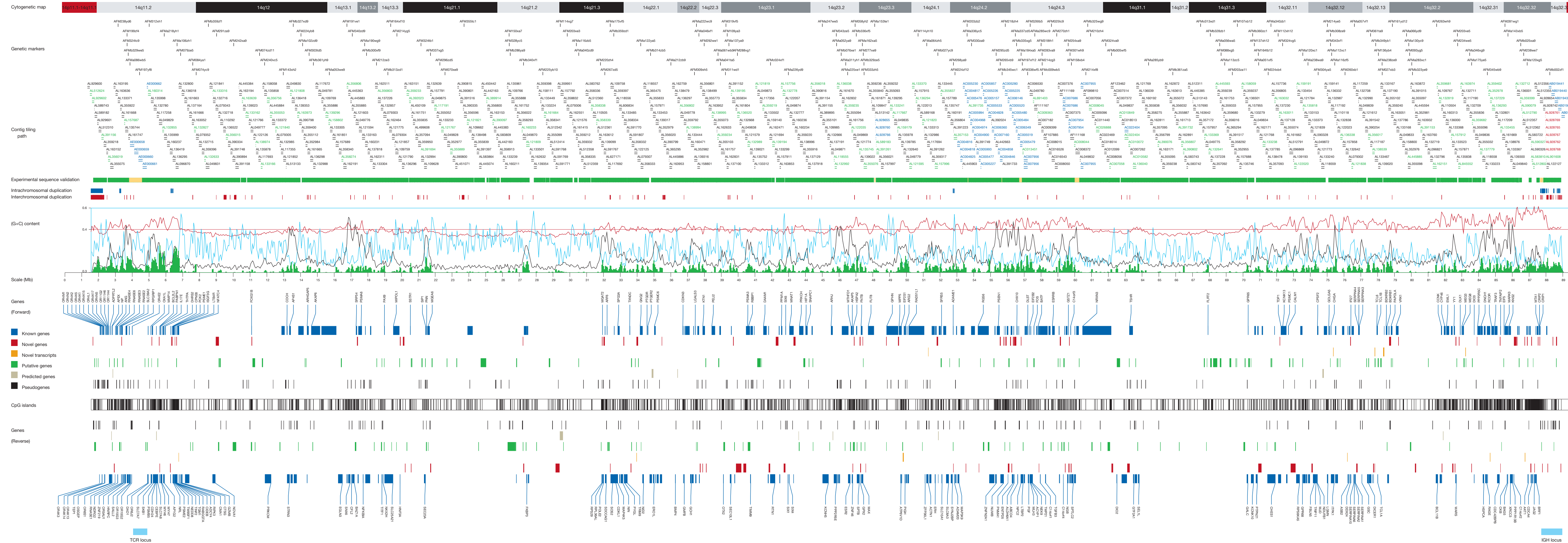
17. Bailey, J. A. *et al.* Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22. *Am. J. Hum. Genet.* **70**, 83–100 (2002).
18. Rowen, L. *et al.* Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity. *Genomics* **79**, 587–597 (2002).
19. Zhen, Y. Y., Libotte, T., Munck, M., Noegel, A. A. & Korenbaum, E. NUANCE, a giant protein connecting the nucleus and actin cytoskeleton. *J. Cell Sci.* **115**, 3207–3222 (2002).
20. Chu, F. F. *et al.* Polymorphism and chromosomal localization of the Gl-form of human glutathione peroxidase (GPX2) on 14q24.1 by *in situ* hybridization. *Genomics* **32**, 272–276 (1996).
21. Araki, O. *et al.* Assignment of type II iodothyronine deiodinase gene (DIO2) to human chromosome band 14q24.2 $\rightarrow$ q24.3 by *in situ* hybridization. *Cytogenet. Cell Genet.* **84**, 73–74 (1999).
22. Hernandez, A., Park, J. P., Lyon, G. J., Mohandas, T. K. & St Germain, D. L. Localization of the type 3 iodothyronine deiodinase (DIO3) gene to human chromosome 14q32 and mouse chromosome 12F1. *Genomics* **53**, 119–121 (1998).
23. Charlier, C. *et al.* Human-ovine comparative sequencing of a 250-kb imprinted domain encompassing the callipyge (clpg) locus and identification of six imprinted transcripts: DLK1, DAT, GTL2, PEG11, antiPEG11, and MEG8. *Genome Res.* **11**, 850–862 (2001).
24. Erdmann, V. A., Szymanski, M., Hochberg, A., de Groot, N. & Barciszewski, J. Collection of mRNA-like non-coding RNAs. *Nucleic Acids Res.* **27**, 192–195 (1999).
25. Modrek, B. & Lee, C. A genomic view of alternative splicing. *Nature Genet.* **30**, 13–19 (2002).
26. Antequera, F. & Bird, A. CpG islands as genomic footprints of promoters that are associated with replication origins. *Curr. Biol.* **9**, R661–R667 (1999).
27. Roest Crolius, H. *et al.* Estimate of human gene number provided by genome-wide analysis using *Tetraodon nigroviridis* DNA sequence. *Nature Genet.* **25**, 235–238 (2000).
28. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).
29. Salamov, A. A. & Solovyev, V. V. *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res.* **10**, 516–522 (2000).
30. Zhao, X. *et al.* Mutations in a newly identified GTPase gene cause autosomal dominant hereditary spastic paraplegia. *Nature Genet.* **29**, 326–331 (2001).
31. Reiss, J. *et al.* A mutation in the gene for the neurotransmitter receptor-clustering protein gephyrin causes a novel form of molybdenum cofactor deficiency. *Am. J. Hum. Genet.* **68**, 208–213 (2001).
32. Stockton, D. W., Das, P., Goldenberg, M., D'Souza, R. N. & Patel, P. I. Mutation of PAX9 is associated with oligodontia. *Nature Genet.* **24**, 18–19 (2000).
33. Dryja, T. P. *et al.* Null RPGRIP1 alleles in patients with Leber congenital amaurosis. *Am. J. Hum. Genet.* **68**, 1295–1298 (2001).
34. Ferda Percin, E. *et al.* Human microphthalmia associated with mutations in the retinal homeobox gene CHX10. *Nature Genet.* **25**, 397–401 (2000).
35. Naureckiene, S. *et al.* Identification of HE1 as the second gene of Niemann-Pick C disease. *Science* **290**, 2298–2301 (2000).
36. Kamnasaran, D. & Cox, D. W. Current status of human chromosome 14. *J. Med. Genet.* **39**, 81–90 (2002).
37. Florea, L., Hartzell, G., Zhang, Z., Rubin, G. M. & Miller, W. A computer program for aligning a cDNA sequence with a genomic DNA sequence. *Genome Res.* **8**, 967–974 (1998).
38. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
39. Shoemaker, D. D. *et al.* Experimental annotation of the human genome using microarray technology. *Nature* **409**, 922–927 (2001).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements We thank H. Riethman, M. Meugnier, C. Sarlande, B. Baude and D. Le Paslier for their respective contributions. We also thank HUGO Gene Nomenclature Committee (H. Wain, R. Laverling, E. Bruford, M. Lush, M. Wright and S. Povey) for determining the chromosome 14 gene symbols.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.H. (e-mail: heilig@genoscope.cns.fr). The entire chromosome 14q sequence is deposited in EMBL under accession number AL954800.



Disruption of fragmented parent bodies as the origin of asteroid families

Patrick Michel*, Willy Benz† & Derek C. Richardson‡

* Observatoire de la Côte d'Azur, BP 4229, 06304 Nice cedex 4, France

† Physikalisches Institut, University of Bern, Sidlerstrasse 5, CH-3012 Bern, Switzerland

‡ Department of Astronomy, University of Maryland, College Park, Maryland 20742-2421, USA

Asteroid families are groups of small bodies that share certain orbit¹ and spectral properties². More than 20 families have now been identified, each believed to have resulted from the collisional break-up of a large parent body³ in a regime where gravity controls the outcome of the collision more than the material strength of the rock. The size and velocity distributions of the family members provide important constraints for testing our understanding of the break-up process, but erosion and dynamical diffusion of the orbits over time can erase the original signature of the collision^{4,5}. The recently identified young Karin family⁶ provides a unique opportunity to study a collisional outcome almost unaffected by orbit evolution. Here we report numerical simulations modelling classes of collisions that reproduce the main characteristics of the Karin family. The sensitivity of the outcome of the collision to the internal structure of the parent body allows us to show that the family must have originated from the break-up of a pre-fragmented parent body, and that all large family members formed by the gravitational reaccumulation of smaller bodies. We argue that most of the identified asteroid families are likely to have had a similar history.

Using a method and numerical codes that have already allowed us to simulate the formation of major asteroid families in different impact energy regimes^{7,8}, we have performed several simulations of the Karin family formation. In such events, the parent body is first totally shattered by the shock wave produced by a colliding projectile into very small fragments, which are then ejected into space. Subsequently, as a consequence of their gravitational interactions, many of the fragments reaccumulate, forming the large members of the family.

To study how the collisional outcome depends on the internal structure of the parent body, we considered different models of such a body and searched for the one which, once disrupted, best matches the observed properties of the Karin family. In particular, we considered two kinds of parent bodies, both spherical in shape:

(1) purely monolithic, and (2) pre-fragmented (pre-shattered or rubble pile). The motivation for modelling the latter kind comes from previous studies of the collisional evolution of the asteroid belt⁹, which suggest that before being dispersed in a large-scale collision, large bodies are likely to have suffered smaller shattering impacts that modified their internal structure. Whereas the internal structure of a monolithic parent body is totally homogeneous, that of a pre-fragmented body is characterized by the presence of fractured zones with no tensile strength. As we shall see, the heterogeneities introduced in this way result in significant changes in the size spectrum of the separate bodies (also called fragments) left after the collision, even though the targets in both cases are completely shattered by the impact. Both types of parent bodies were modelled with material properties corresponding to either basalt or dunite, using the Tillotson¹⁰ or ANEOS¹¹ equations of state, respectively. The corresponding reference bulk densities of these materials are 2.7 g cm^{-3} for basalt and 3.33 g cm^{-3} for dunite, which are within the range of estimated bulk densities of S-type asteroids^{12,13} and ordinary chondrites¹⁴.

The diameter of the Karin parent body was inferred from the estimated sizes of the observed members of the family plus a reasonable estimate of the observational incompleteness of the main belt for asteroids with absolute magnitude $H < 16.0$ to account for undetected members. This method gives a parent body diameter D of about $24.5 \pm 1 \text{ km}$, the largest remnant being estimated to account for $47 \pm 6\%$ of its mass⁶.

In our simulations (see Table 1), we consider a parent body 25 km in diameter represented by 2×10^5 particles. For each model of internal structure, we consider at most two different impact geometries and adjust the projectile's size to yield a largest remnant with mass in agreement with that inferred from observations. In all cases but one, the impact velocity is fixed at 5 km s^{-1} , which corresponds to the average collisional speed in the asteroid belt¹⁵.

In all our simulations, the fragmentation phase driven by the shock wave of the impact and computed using a three-dimensional smoothed particle hydrodynamics (SPH) code¹⁶, leads to the complete pulverization of the parent body down to the resolution limit, which corresponds to a fragment radius around 0.2 km. The subsequent evolution, including fragment collisions and reaccumulations, is then computed with a sophisticated N -body code¹⁷ until the outcome reaches a nearly steady state (typically a few days of simulated real time). At this point, gravitational reaccumulation has led to the formation of a full size spectrum of aggregates (also called fragments), whose properties we compare with the real Karin cluster. For this, we use two different diagnostics: (1) the cumulative size distribution of the fragments; and (2) their ejection velocities, or equivalently, the dispersion of the family in proper orbital element space. However, any attempt to match the observed proper-

Table 1 Characteristics of the Karin family formation simulations

Parent body structure	Equation of state	R_{proj} (km)	V_{proj} (km s ⁻¹)	θ (°)	Q (erg g ⁻¹)	$M_{\text{lr}}/M_{\text{pb}}$	V_{lr} (m s ⁻¹)	$\langle V_{\text{ej}} \rangle$ (m s ⁻¹)
Monolithic (basalt)	T	1.35	5	0	1.57×10^8	0.52	5.8	18
Monolithic (dunite)	A	1.22	5	0	1.16×10^8	0.54	0.4	19
Monolithic (dunite)	A	3.10	1	0	7.61×10^7	0.50	0.9	18
Monolithic (basalt)	T	1.65	5	45	2.87×10^8	0.50	8.8	19
Monolithic (dunite)	A	1.57	5	45	2.47×10^8	0.52	5.6	14
Pre-shattered (basalt)	T	1.03	5	0	6.97×10^7	0.47	3.7	17
Pre-shattered (dunite)	A	1.13	5	0	9.21×10^7	0.52	3.6	14
Pre-shattered (basalt)	T	1.21	5	45	1.13×10^8	0.51	4.2	15
Pre-shattered (dunite)	A	1.45	5	45	1.95×10^8	0.46	5.0	18
Rubble pile (basalt)	T	1.50	5	45	2.16×10^8	0.49	5.0	14

Three types of parent bodies were considered, differing in their internal structure: monolithic, pre-shattered or rubble pile. To model a pre-shattered parent body, we generated a network of cracks in an originally monolithic body, dividing it into 40 distinct fragments distributed uniformly in radius and separated by fractured zones. Fragment masses were randomly distributed in a range with the maximum mass four times larger than the minimum mass. Experiments with other position and mass distributions produced similar results. To model a rubble pile, we generated solid boulders whose centres were chosen at random inside the parent body, with their individual radii computed from a power-law distribution. In most cases two simulations were performed for each type of parent body, one using material properties of basalt and the Tillotson equation of state (T) and one using dunite and the ANEOS equation of state (A). Different impact conditions were also used, defined by the projectile's radius R_{proj} , impact velocity V_{proj} and angle of incidence θ . The specific impact energy is given by $Q = (\text{projectile kinetic energy})/(\text{target mass})$. $M_{\text{lr}}/M_{\text{pb}}$ is the resulting mass ratio of the largest remnant to the parent body and V_{lr} is the largest remnant ejection speed. The mean ejection speed of all the fragments is given by $\langle V_{\text{ej}} \rangle$.

ties exactly is pointless given the uncertainties in the observational data (for example, sizes are only estimated using inferred albedos), as well as the large number of unknowns characterizing the impact (actual impact velocity, shapes, and so on). That is why we focus more on global characteristics that we believe to be robust, such as the continuous aspect of the cumulative size distribution and the shape of the cluster in proper element space.

In the case of the break-up of monolithic parent body, we find that the cumulative size distribution of fragments is always characterized by a lack of intermediate-sized bodies and a very steep slope for the smaller ones. This trend also holds true for a collision occurring at a smaller (and much less likely) impact speed of 1 km s^{-1} but with a bigger projectile to yield a largest remnant of similar mass (see Fig. 1a, b). Yet the inferred size distribution of the real Karin family is rather continuous (the second-largest body is estimated to have a diameter $D \approx 14.4 \text{ km}$).

Disruptions involving a pre-fragmented (pre-shattered or rubble pile) parent body, on the other hand, result in a much more

continuous cumulative size distribution of fragments (Fig. 1c, d). Hence the presence of intermediate-sized bodies, as observed in the real family, is most probably a direct consequence of the presence of large-scale fractures or big blocks within the parent body. Such large members cannot be obtained starting with a monolithic parent body regardless of the impact geometry and material type in all the cases investigated (not all are shown here), so we conclude that the parent body of the Karin family must have been pre-fractured or reaccumulated. This is consistent with the fact that the Karin parent body actually belonged to the older Koronis family, for which we have shown in previous simulations that the members down to at least kilometre sizes are reaccumulated objects^{7,8,16}.

Interestingly, we note that intermediate-size fragments are also present in most major asteroid families, which implies that many parent bodies in the asteroid belt were probably pre-fragmented. Consequently, this must also be the case for most large objects in the present-day asteroid belt, which is consistent with the idea that asteroids get battered over time in the general sense.

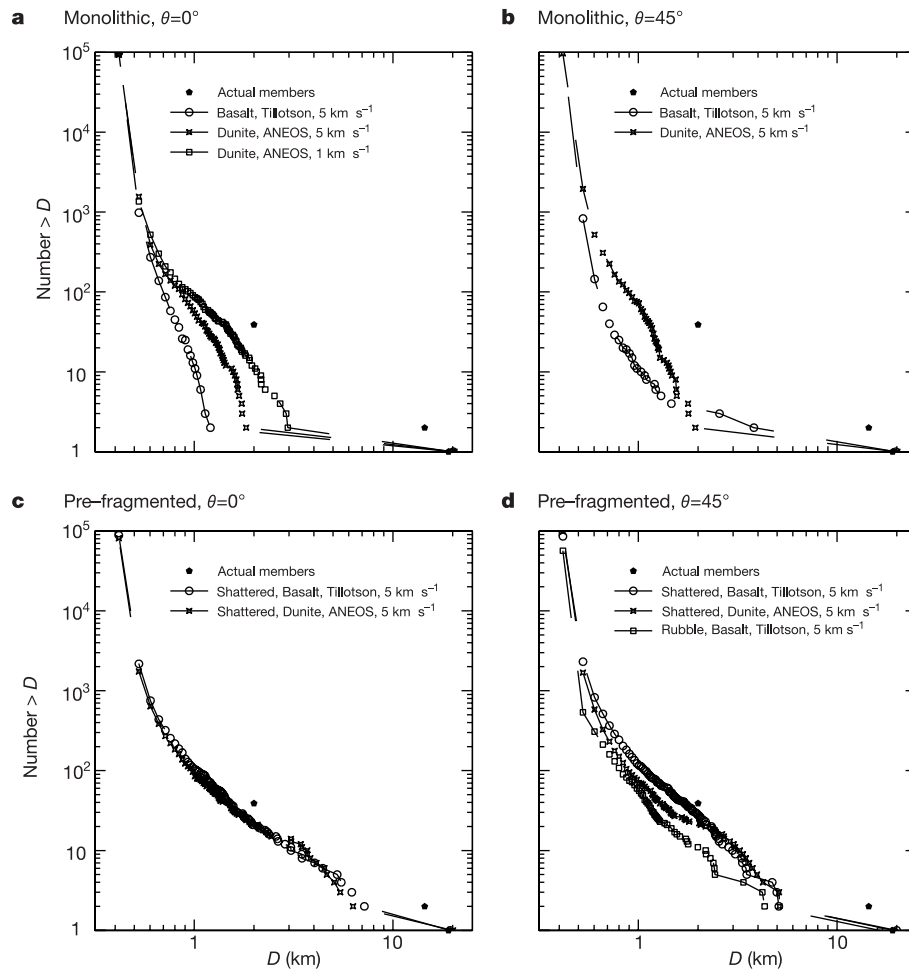


Figure 1 Cumulative diameter distributions of the fragments of the simulated Karin families. **a, b**, The disruption of a monolithic parent body; **c, d**, the results for a pre-fragmented parent body. **a** and **c** were obtained with a projectile colliding 'head on', whereas an impact with an angle of incidence θ equal to 45° produced **b** and **d**. Symbols denote the different material properties or impact parameters used in the simulations, as indicated on the plots. In all simulations, the parent body had a diameter D equal to 25 km . The plots also show the estimated sizes of the two largest members as well as the number of members larger than 2.1 km (computed from their absolute magnitudes and assuming a 16% albedo for all family members; see ref. 6). The best agreement is obtained with

our simulations of the disruption of a pre-fragmented parent body and does not depend much on the impact geometry or the material type and equation of state. Conversely, in the case of a monolithic target, different impact geometries and/or material types and equations of state do not produce more intermediate-sized aggregates, even though some quantitative differences exist. For instance, a 45° impact results in a somewhat smaller gap between the sizes of the largest and the second-largest remnant in the simulation of a basaltic parent body with a Tillotson equation of state. We note that only the simulations involving a pre-fragmented parent body yield continuous size distributions.

The second important constraint we considered was the ejection velocity distribution of simulated fragments, which must be compatible with that of the real family members. Karin's young age (around 5 Myr) implies that dynamical diffusion has had no time to alter appreciably the initial ejection velocity distribution, which is not the case for the other, much older, families. Unfortunately, only the proper orbital elements (semi-major axis a , eccentricity e , and inclination I) of the family members can be deduced from observations, whereas the simulations only provide the ejection velocities. However, Gauss' equations can be used to relate ejection velocities to orbital elements, provided the position of the barycentre of the family in a heliocentric reference frame is known. To solve these equations, two additional unknown angles must be specified, namely the true anomaly f and the argument of perihelion

ω of the parent body's orbit at the time of the impact. The current observed shape of the cluster in orbital element space suggests $f = 30^\circ$ and $f + \omega = 45^\circ$ as a good choice⁶.

As can be seen from Fig. 2, the dispersion in ejection velocities of fragments originating from a monolithic parent body impacted at 5 km s^{-1} is too small to explain the actual observed spread in proper element space. Conversely, supporting our previous conclusion based on the size distribution, the ejection velocity field obtained in a collision involving pre-fragmented parent bodies is in excellent agreement with the observed family. Such agreement is also obtained from the break-up of a monolithic parent body at low impact speed (1 km s^{-1}); however, as remarked before, in this case the fragment size distribution is not continuous. In summary, a good match of both a continuous size distribution and the orbital

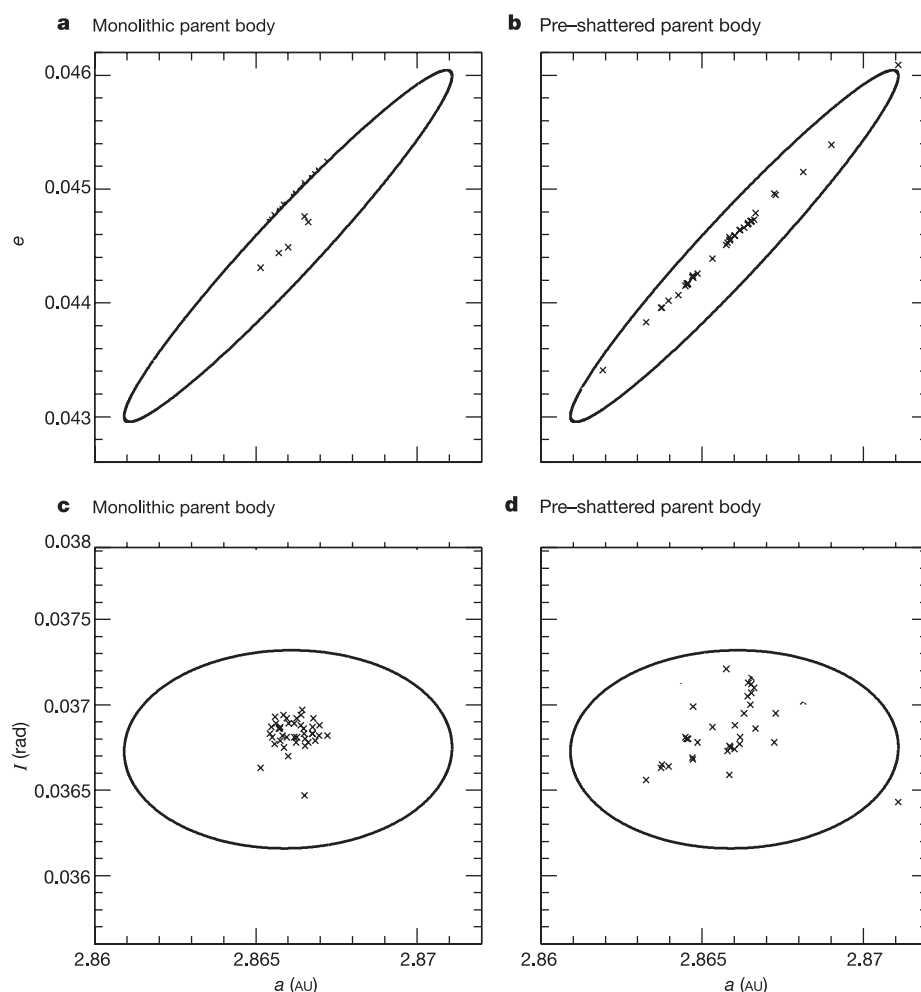


Figure 2 Dispersion of the simulated Karin family in orbital element space. The plots show the distributions in the eccentricity versus semi-major axis plane (**a**, **b**) and in the inclination versus semi-major axis plane (**c**, **d**) obtained by considering either a monolithic parent body (**a**, **c**) or a pre-shattered one (**b**, **d**) of 25 km diameter suffering a 'head-on' collision with a projectile. The ANEOS equation of state was used with material properties of dunite. Because the proper elements of the real family members are not available, and because we are not interested in fitting exactly the position of each real member but rather in reproducing the global dispersion of the Karin family, we show the ellipse already identified⁹ as best matching the dispersion of the real family in this space. The orbital elements of fragments are computed from their ejection velocities using Gauss' equations. To convert ejection velocities from our simulations into orbital elements, we first need to define the orientation of the velocity components of our fragments in a heliocentric frame. For this, the reference frame used in our simulations being arbitrary, we must specify the direction of impact of the projectile. Different directions lead to

different shapes of the family in orbital element space, so we investigated a range of possible orientations which are then obtained by assuming the values of the semi-major axis and inclination of the projectile's orbit, and using the so-called Tisserand constant as a constraint to fix the eccentricity (see ref. 8 for a description of this procedure). We consider here one possible orientation. Other combinations of the projectile's orbital elements or impact direction would not necessarily lead to the same shape of cluster. The values of the parent body's true anomaly at impact f and its sum with the argument of perihelion $\omega + f$ are assumed to be $f = 30^\circ$ and $\omega + f = 45^\circ$; its semi-major axis a , eccentricity e , and inclination I are taken to be $a = 2.866 \text{ AU}$, $e = 0.0445$ and $I = 2.105^\circ$. Only 39 real Karin family members are currently known, so the figure also shows the distributions of only the 39 largest fragments of our simulation, the smallest among them having a size of 1.5 km, to be compared to the estimated 2.1 km for the real one. For the projectile, the velocity of impact is 5 km s^{-1} , $\theta = 0^\circ$, $a = 2.2 \text{ AU}$, and $I = 5^\circ$.

dispersion of the Karin family requires that its parent body was pre-fragmented or reaccumulated. This supports the overall picture that all large members of asteroid families are reaccumulated bodies, given that the Karin cluster is at least a second-generation family in the lineage of the older Koronis parent body.

Our simulations can also be used to determine the impact energy needed to produce a given degree of disruption as a function of the internal structure of the parent body. From Table 1, we see that disrupting a pre-shattered target requires less energy per unit mass than for a monolithic body. This, at first glance a surprising result, is related to the fact that the fractures as modelled here (no porosity and no material discontinuities except damage) do not affect shock waves but only tensile waves. Hence, fragments can be set in motion immediately upon being hit by the shock wave without having to wait for fracture to occur in a following tensile wave. Thus, transfer of momentum is more efficient and disruption facilitated. Note that the presence of large voids such as those in rubble piles affect the propagation of the shock wave, thus reversing this trend. A more detailed study of these properties will be presented in another paper. Nevertheless, it is already clear that the internal structure of an asteroid plays an important role in the determination of its response to impacts (see also ref. 18). This is not only relevant for estimating the collisional lifetime of a body in the asteroid belt, but also for developing strategies to deflect a potential Earth impactor.

A last important implication of our simulations concerns the properties of family members with sizes smaller than can currently be observed. The break-up of a pre-shattered Karin parent body produces several thousand fragments with sizes ranging from a few kilometres down to our resolution limit of 0.2 km (Fig. 1). Their ejection speeds can allow some of them to be injected into efficient transport mechanisms leading to Earth-crossing orbits. Therefore, we conclude that the Karin break-up event may well have produced some near-Earth asteroids and even some meteoroids. According to our simulations, the potential near-Earth asteroids with size at least above 0.2 km result from the gravitational reaccumulation of smaller fragments, and are therefore all aggregates. □

Received 4 October; accepted 5 December 2002; doi:10.1038/nature01364.

1. Milani, A. & Knežević, Z. Asteroid proper elements and the dynamical structure of the asteroid main belt. *Icarus* **107**, 219–254 (1994).
2. Zappalà, V., Bendjoya, P., Cellino, A., Farinella, P. & Froeschlé, C. Asteroid families: Search of a 12,487-asteroid sample using two different clustering techniques. *Icarus* **116**, 291–314 (1995).
3. Farinella, P., Davis, D. R. & Marzari, F. in *Completing the Inventory of the Solar System* (eds Rettig, T. W. & Hahn, J. M.) 45–55 (ASP Conference Series 107, Astronomical Society of the Pacific, San Francisco, 1996).
4. Marzari, F., Farinella, P. & Davis, D. R. Origin, aging, and death of asteroid families. *Icarus* **142**, 63–77 (1999).
5. Bottke, W. F., Vokrouhlický, D., Brož, M., Nesvorný, D. & Morbidelli, A. Dynamical spreading of asteroid families via the Yarkovsky effect. *Science* **294**, 1693–1696 (2001).
6. Nesvorný, D., Bottke, W. F., Dones, L. & Levison, H. F. The recent breakup of an asteroid in the main-belt region. *Nature* **417**, 720–722 (2002).
7. Michel, P., Benz, W., Tanga, P. & Richardson, D. C. Collisions and gravitational reaccumulation: Forming asteroid families and satellites. *Science* **294**, 1696–1700 (2001).
8. Michel, P., Benz, W., Tanga, P. & Richardson, D. C. Formation of asteroid families by catastrophic disruption: Simulations with fragmentation and gravitational reaccumulation. *Icarus* **160**, 10–23 (2002).
9. Melosh, J. G. & Ryan, E. V. Asteroids: shattered but not dispersed. *Icarus* **129**, 562–564 (1997).
10. Tillotson, J. H. *Metallic Equations of State for Hypervelocity Impact* (General Atomic Report GA-3216, San Diego, 1962).
11. Thompson, S. L. & Lauson, H. F. Improvement in the chart D radiation hydrodynamic code III: revised analytic equation of state. (Report SC-RR-71 0714, Sandia National Laboratory, Albuquerque, 1972).
12. Belton, B. *et al.* The bulk density of asteroid 243 Ida from Dactyl's orbit. *Nature* **374**, 785–788 (1995).
13. Yeomans, D. K. *et al.* Radio science results during the NEAR-Shoemaker spacecraft rendezvous with Eros. *Science* **289**, 2085–2088 (2000).
14. Flynn, G. J., Moore, L. B. & Klöck, W. Density and porosity of stone meteorites: implications for the density, porosity, cratering, and collisional disruption of asteroids. *Icarus* **142**, 97–105 (1999).
15. Bottke, W. F., Nolan, M. C., Greenberg, R. & Kolvoord, R. A. Velocity distributions among colliding asteroids. *Icarus* **107**, 255–268 (1994).
16. Benz, W. & Asphaug, E. Catastrophic disruptions revisited. *Icarus* **142**, 5–20 (1999).
17. Richardson, D. C., Quinn, T., Stadel, J. & Lake, G. Direct large-scale *N*-body simulations of planetesimal dynamics. *Icarus* **143**, 45–59 (2000).
18. Asphaug, E. *et al.* Disruption of kilometre-sized asteroids by energetic collisions. *Nature* **393**, 437–440 (1998).

Acknowledgements P.M. acknowledges financial support from the Action Thématique Innovante 2001 of the French INSU, the programme 'Bonus-Qualité-Recherches 2001' and the Cassini laboratory of the Côte d'Azur Observatory (OCA). W.B. and D.C.R. acknowledge support respectively from the Swiss National Science Foundation and NASA through the Office of Space Science. Simulations were carried out on 4-processor Compaq DEC Alpha workstations thanks to the SIVAM project and the ILGA team of the OCA and on a Beowulf installed by the society Alineos.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.M. (e-mail: michel@obs-nice.fr).

Attosecond control of electronic processes by intense light fields

A. Baltuška[†], Th. Udem[†], M. Uiberacker^{*}, M. Hentschel^{*}, E. Goulielmakis[†], Ch. Gohle[†], R. Holzwarth[†], V. S. Yakovlev^{*}, A. Scrinzi^{*}, T. W. Hänsch[†] & F. Krausz^{*}

^{*} Institut für Photonik, Technische Universität Wien, Gusshausstrasse. 27, A-1040 Wien, Austria

[†] Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Strasse 1, D-85748 Garching, Germany

The amplitude and frequency of laser light can be routinely measured and controlled on a femtosecond (10^{-15} s) timescale¹. However, in pulses comprising just a few wave cycles, the amplitude envelope and carrier frequency are not sufficient to characterize and control laser radiation, because evolution of the light field is also influenced by a shift of the carrier wave with respect to the pulse peak². This so-called carrier-envelope phase has been predicted^{3–9} and observed¹⁰ to affect strong-field phenomena, but random shot-to-shot shifts have prevented the reproducible guiding of atomic processes using the electric field of light. Here we report the generation of intense, few-cycle laser pulses with a stable carrier envelope phase that permit the triggering and steering of microscopic motion with an ultimate precision limited only by quantum mechanical uncertainty. Using these reproducible light waveforms, we create light-induced atomic currents in ionized matter; the motion of the electronic wave packets can be controlled on timescales shorter than 250 attoseconds (250×10^{-18} s). This enables us to control the attosecond temporal structure of coherent soft X-ray emission produced by the atomic currents—these X-ray photons provide a sensitive and intuitive tool for determining the carrier-envelope phase.

Matter exposed to intense laser light undergoes ionization, which gives rise to a broad range of phenomena in atoms, molecules and plasmas. An electronic wave packet is set free around each oscillation peak of a laser electric field that is strong enough to overcome the effective binding potential (Fig. 1a). The ensuing motion of the wave packets released by optical-field ionization depends on the subsequent evolution of the driving laser field^{11,12}. A laser pulse consisting of many wave cycles launches a number of wave packets at different instants. Each of these follows a different trajectory because of the differences in the initial conditions of their motion, preventing a precise control of strong-field-induced electronic dynamics. Intense few-cycle light pulses with adjustable carrier-envelope (C-E) phase hold promise for a marked improvement. With the C-E phase set to make the peak of the oscillating electric field coincide with the pulse peak (blue line in Fig. 1c) and the strength of the field just sufficient to reach the ionization threshold

at the pulse centre, a single, isolated electronic wave packet can be formed. The motion of this wave packet launched into an accurately controlled electric (and magnetic) field at a well-defined instant is determined with the highest possible precision permitted by quantum mechanics. Hence, few-cycle light with controlled electric-field waveforms offers the ultimate control of strong-field-driven atomic, molecular and plasma dynamics. This prospect constitutes the primary motivation for the work reported here.

In the present experiments, we adjust the peak intensity of linearly polarized few-cycle laser pulses such that several electronic wave packets in the vicinity of the pulse peak are set free. First they are removed from their parent ion, but within a laser period they are pulled back by the laser electric field (Fig. 1b). The highest-energy portion of the wave packet re-collides with the ion near the second (counting from the wave packet's 'birth') zero transition of the laser electric field, and results in the emission of an energetic (soft-X-ray) photon^{11,13}. Being repeated many times in a multi-cycle laser field, this elementary process results in a series of equidistant emission

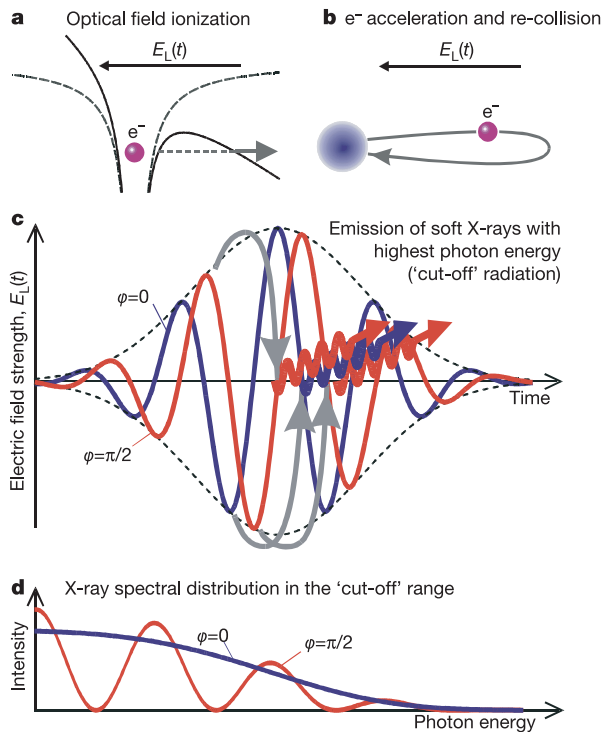


Figure 1 Optical-field ionization and generation of coherent extreme ultraviolet and soft-X-ray radiation from an atom exposed to a strong, linearly polarized, few-cycle light pulse. **a**, The effective Coulomb potential binding valence electrons to the atomic core (dashed curve) is temporarily suppressed around the oscillation peak of the laser electric field. A valence electron can tunnel through or escape above the potential barrier formed by the superposition of the atomic Coulomb field and the instantaneous laser field (solid curve). The electron wave packet that is set free within one-half of the laser period ($T_0 = 2.5$ fs at a wavelength of 750 nm) is temporally confined to less than $T_0/10$ owing to a nonlinear dependence of the ionization probability on the strength of the laser electric field E_L . Atoms exposed to a few-cycle pulse emit one or several wave packets near the pulse peak, depending on the peak intensity (relative to the ionization threshold) and the C-E phase φ . **b**, The freed electron is moved away from the atomic core and then pulled back to it by a linearly polarized field. Re-collision of the electron with its parent ion may trigger several processes, including secondary electron emission, excitation of bound electrons and emission of an energetic (soft-X-ray) photon. **c**, The highest-energy ('cut-off') X-ray photons are emitted near the zero transitions of the laser electric field around the pulse peak. With $\varphi \approx 0$ (cosine waveform, blue line) they are temporally confined to one single burst, whereas $\varphi \approx \pi/2$ (sine waveform, red line) yields two bursts of comparable amplitude separated by $\sim T_0/2$. **d**, As a result, the spectral distribution of the emitted 'cut-off' X-rays is continuous or modulated quasi-periodically, respectively.

lines corresponding to high-order odd harmonics of the driver laser light¹⁴ and a sequence of sub-femtosecond bursts in the time domain¹⁵. By contrast, our few-cycle pulses release only a few wave packets near the pulse centre. The grey arrows in Fig. 1c depict

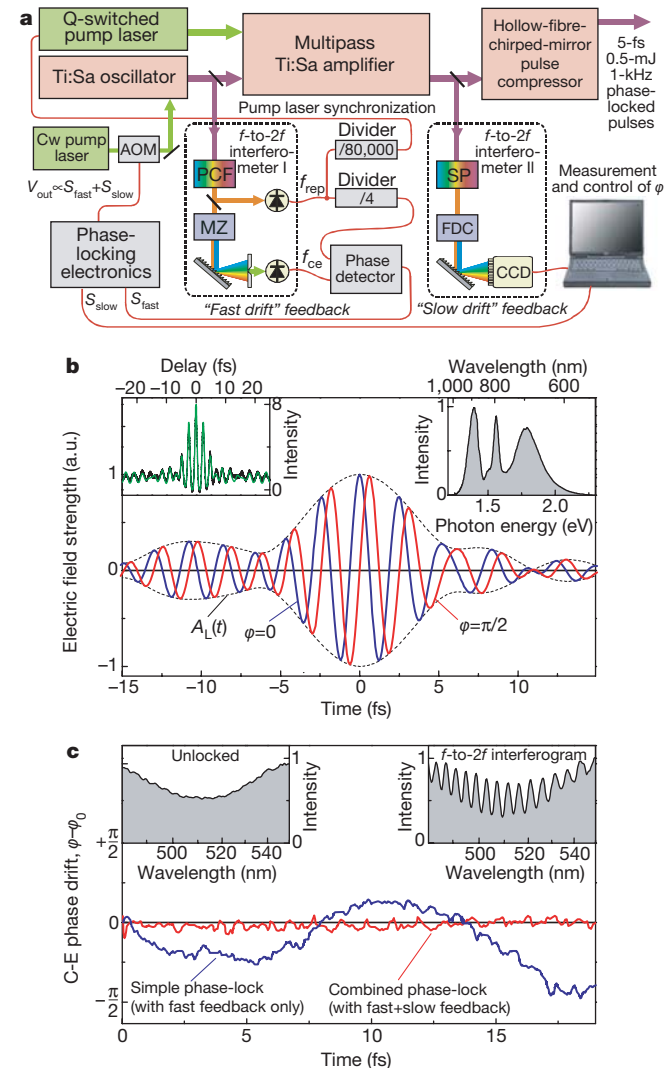


Figure 2 Overview of C-E-phase-stabilized high-power laser system. **a**, Schematic of the laser set-up. AOM, acousto-optical modulator; PCF, photonic crystal fibre; MZ, Mach-Zehnder set-up; SP, 2-mm sapphire plate; FDC, frequency-doubling crystal. The C-E phase of the pulses delivered by the titanium:sapphire (Ti:Sa) oscillator is controlled by tracking the f -to- $2f$ signal in interferometer I and controlling the pump power (Verdi, Coherent) through a feedback based on the AOM. Frequency dividers $/4$ and $/80,000$ are used to derive, respectively, the reference frequency for the stabilization of the phase slip behind interferometer I and the repetition rate of pulses amplified in a multipass amplifier (Femtopower Compact Pro, Femtolasers; pump source: Corona, Coherent). With this technique the C-E phase recurrence of the Ti:Sa oscillator is set such that every fourth pulse (and therefore also every 80,000th pulse) exhibits the same electric-field waveform. The residual drift of the C-E phase behind the laser amplifier is monitored with interferometer II and pre-compensated by shifting accordingly the C-E phase of the oscillator. See Methods for further details. **b**, Pulse properties behind the hollow-fibre-chirped-mirror compressor. The blue and red curves depict possible electric-field waveforms compatible with $A_L(t)$ and $\omega_L(t)$ retrieved from the interferometric autocorrelation (left inset) and the pulse spectrum (right inset) for different settings of the C-E phase; **c**, C-E phase drift versus time with and without slow-drift feedback (compare **a**); insets show the f -to- $2f$ interference spectrograms recorded with interferometer II. a.u., arbitrary units; f_{rep} , repetition rate of the oscillator; f_{ce} , f -to- $2f$ beat signal; S_{fast} and S_{slow} , electric signals from the 'fast' and 'slow' feedback loops; V_{out} , amplitude of AOM control voltage.

those that have the highest kinetic energy at re-collision and, consequently, produce the highest-energy (cut-off) X-ray photons. The spectral distribution of these photons depends sensitively on whether they are emitted in a single sub-femtosecond X-ray burst or in a couple of bursts, which can be controlled with the C-E phase of the few-cycle laser pulse (Fig. 1d). Such sub-femtosecond X-ray bursts were recently produced and measured using randomly phased few-cycle light pulses^{16,17}. As the emission of two bursts of comparable energy can only occur within a narrow range of C-E phases, statistically, the highest-energy X-ray photons were carried in one dominant burst. Yet, optimization of the excitation conditions for maximum concentration of X-ray emission energy in a single pulse and their accurate reproduction has not been possible with randomly phased laser pulses. Here we report the generation of this radiation—which we do not refer to as high harmonics for the

reasons explained below—with phase-stabilized light pulses. The spectral distribution of the highest-energy X-ray photons emerging from the few-cycle-driven ionization process enables us to determine the C-E phase. In combination with standard pulse diagnostics, this approach permits complete characterization of the temporal structure of few-cycle light waveforms (with an ambiguity in the direction of the field).

The electric field of a linearly polarized light pulse can be generally written as $E_L(t) = A_L(t) \cos[\omega_L(t)t + \varphi]$. This description includes three physical quantities: the amplitude envelope $A_L(t)$, where t is time, the instantaneous frequency of the field oscillations $\omega_L(t) = \omega_0 + \beta(t)$, where $\beta(t)$ is a possible frequency sweep across the pulse, and the C-E phase φ . In few-cycle pulses, a change in φ affects $E_L(t)$ considerably². Whereas $A_L(t)$ and $\omega_L(t)$ can be obtained from standard pulse diagnostic techniques and flexibly modified in pulse shapers, φ has remained immeasurable so far. In general, pulses delivered by femtosecond lasers have a reproducible envelope and carrier frequency but carry a random C-E phase². It was demonstrated recently that the pulse-to-pulse shift of φ can be tracked in real time and stabilized using a so-called f -to- $2f$ interferometer^{18–21}. In such a scheme, the laser spectrum is broadened to a full octave and the beat between the frequency-doubled low-frequency spectral components and the high-frequency components is detected and used for active phase stabilization.

In our experiments, we amplify phase-stabilized pulses from a sub-10-fs titanium:sapphire (Ti:Sa) laser (seed oscillator) in a multipass chirped-pulse Ti:Sa amplifier operating at a 1-kHz repetition rate (Fig. 2a). Amplification and temporal recompression yield 20-fs, 1-mJ pulses, which are subsequently spectrally broadened by self-phase modulation in a hollow-core waveguide filled with neon gas, and finally compressed on reflection off chirped multilayer mirrors. This system delivers 5-fs, 0.5-mJ pulses with a carrier wavelength of 750 nm (Fig. 2b)²². The C-E phase jitter emerging in the waveguide has been measured by comparing the output beam with a reference split off the input beam in a linear interferometer. The jitter is found to be less than 50 mrad (root-mean-square, r.m.s.) owing to the excellent energy stability of the amplified pulses (<1% r.m.s.). The C-E phase can therefore be monitored directly at the output of the amplifier with a broadband f -to- $2f$ interferometer^{23,24}.

The drift of the C-E phase can be retrieved from the spectral shift of the recorded pattern of the f -to- $2f$ spectral interference (right inset in Fig. 2c). In the absence of phase stabilization of the seed oscillator, the f -to- $2f$ interferogram is completely blurred after approximately 200 laser shots (left inset of Fig. 2b). In contrast, seeding the amplifier with phase-stabilized pulses preserves good fringe visibility over thousands of laser shots. The phase evolution derived from the change of the interference pattern (blue curve in Fig. 2c) appears to be frozen for periods as long as several seconds. Considering the fact that some extra phase noise might be introduced in the measurement process, the blue curve in Fig. 2c accounts for the maximal possible drift of the C-E phase present in the laser system. Due to the slowness of this drift, it can be readily tracked by computer analysis of the f -to- $2f$ interferogram, and pre-compensated by a phase offset of opposite sign through a ‘slow-drift’ feedback loop (Fig. 2a). The simultaneous use of both feedback loops allows stabilizing the C-E phase for extended periods ($\gg 1$ minute). The residual variation of the C-E phase retrieved using the in-loop f -to- $2f$ interferometer is shown by red curve in Fig. 2c. In this case, the upper limit of the C-E phase drift at the output of the laser system can be obtained with an additional, out-of-loop interferometer.

With these intense phase-controlled few-cycle pulses, we can now set out to control electronic motion within the field oscillation cycle of visible light and exploit this technical capability to measure the value of the C-E phase by gauging it with a sensitive strong-field interaction. According to the intuitive insight into the underlying

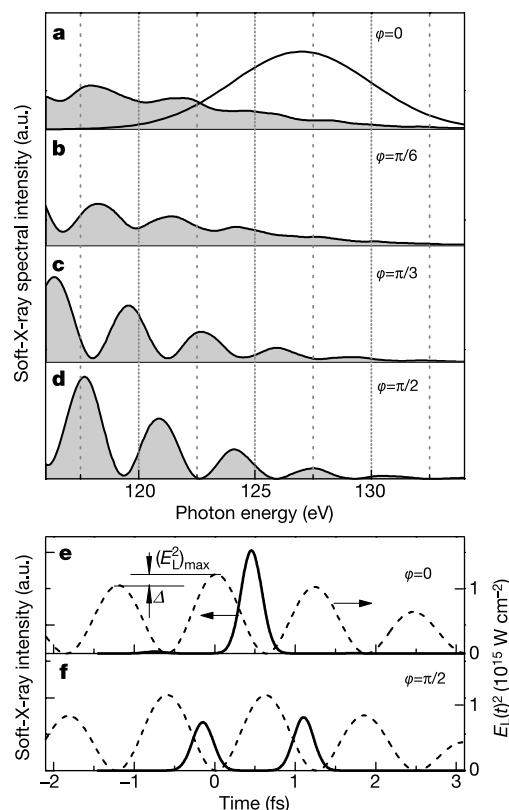


Figure 3 Numerical simulations of few-cycle-driven coherent soft-X-ray emission from ionizing atoms. **a–d**, Cut-off-range spectra for different C-E-phase settings of the driving laser pulse. **e, f**, The solid curves depict the temporal intensity profile of the cut-off harmonic radiation filtered through a Gaussian bandpass filter with a full-width at half-maximum of 7 eV (solid line in **a**), whereas the dashed curves plot $E_L(t)^2$. The parameters of the calculations are as follows: laser pulse duration, $\tau_L = 5$ fs; pulse energy, 0.2 mJ; beam diameter, $2w_L = 122 \mu\text{m}$; medium, neon; pressure, 100 mbar; length, 2 mm. These results, along with previous numerical studies^{4,9}, fully support our intuitive analysis: the few-cycle cosine wave is predicted to produce a single soft-X-ray burst (filtered in the cut-off, **e**). Deviation of φ from zero gradually suppresses the magnitude of the main burst and gives rise to a satellite spike. The latter becomes most prominent for $|\varphi| \rightarrow \pi/2$ (**f**). In the frequency domain, the isolated pulse emerging for $\varphi = 0$ implies a continuous spectrum (**a**), which gets increasingly modulated with the appearance of the second burst for $|\varphi| \rightarrow \pi/2$ (**b–d**). Detailed numerical simulations show that the values of φ corresponding to the smoothest and the most deeply modulated cut-off spectra are shifted by about 20° from the intuitively anticipated values of 0 and $\pi/2$. The relative bandwidth of the cut-off continuum linearly scales with the ratio $\Delta/(E_L^2)_{\text{max}}$ (compare **e**), which, in turn, is related to the number of cycles in the optical field, τ_L/T_0 . In fact, the continuum bandwidth is as broad as ~ 20 eV for $\tau_L/T_0 = 2$ and a cut-off photon energy of 125 eV, and shrinks to only 3.8 eV for $\tau_L/T_0 = 5$.

physics, outlined above, the energetic radiation from an atom ionized by a linearly polarized few-cycle pulse constitutes an excellent candidate for such a C-E-phase calibration. Owing to the fact that the emission of the highest-energy (cut-off) photons is confined to about a single light oscillation period, T_0 , the time structure of the generated X-rays is expected to depend sensitively on the C-E phase of the few-cycle driver light. In fact, for a cosine waveform featuring a single strongest half-cycle ($\varphi \approx 0 \pm n\pi$, where $n = 0, 1, \dots$), the highest-energy photons are emitted within a single burst resulting from the wave packet that was 'born' near $t = -T_0/2$ and re-collided with the parent ion in the vicinity of $t = T_0/4$ (Fig. 1c). For a sine waveform ($\varphi \approx \pi/2 \pm n\pi$) there are two most intense half cycles resulting in a pair of most energetic re-colliding wave packets and, correspondingly, two cut-off X-ray bursts emitted near $t = 0$ and $T_0/2$ (Fig. 1c). Therefore, for these two different values of φ we expect distinctly different spectral distributions of the highest energy photons (Fig. 1d): a continuous distribution for $\varphi \approx 0 \pm n\pi$ (where $n = 0, 1, \dots$), and a distribution exhibiting deep (quasi-) periodic modulation in the case of $\varphi \approx \pi/2 \pm n\pi$. These intuitive considerations have been fully confirmed by our numerical simulations^{25,26} for the realistic

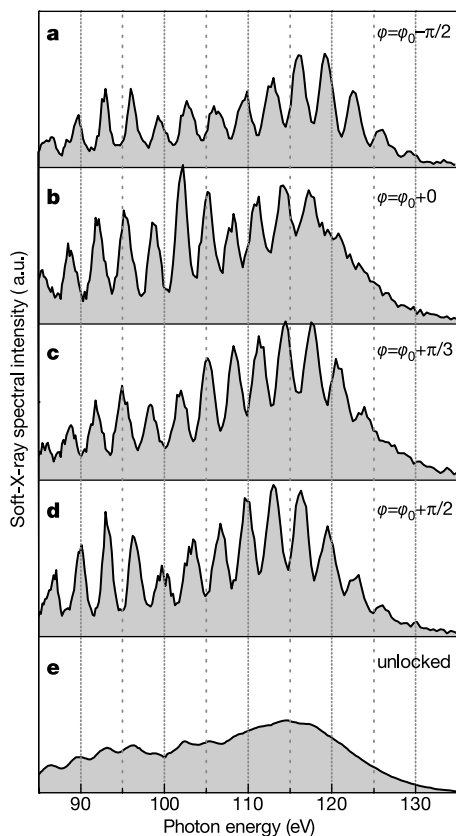


Figure 4 Measured spectral intensity of few-cycle-driven soft-X-ray emission from ionizing atoms. **a–d**, Data obtained with phase-stabilized pulses for different C-E-phase settings. **e**, Spectrum measured without phase stabilization. CCD exposure time was 0.5 s in all cases. The coherent energetic radiation was generated by gently focusing 5-fs, 0.2-mJ laser pulses into a 2-mm-long neon gas medium. The on-axis peak intensity of the pump pulse was estimated to be $7 \times 10^{14} \text{ W cm}^{-2}$. The neon gas was supplied in a thin-walled metal tube with a backing pressure of 160 mbar. The pressure in the interaction region, however, was somewhat lower owing to the gas expansion into the surrounding vacuum chamber. Thin zirconium foils were installed in the pathway of the generated soft-X-ray radiation in order to block the laser light as well as low-order harmonics. The high-energy part of the spectrum (above 80 eV) was spectrally analysed with a simple spectrometer consisting of a $10,000 \text{ lines mm}^{-1}$ transmission grating and a back-illuminated soft-X-ray CCD camera (Roper Scientific).

experimental conditions of a macroscopic generation medium. Figure 3 depicts representative results. Our investigations reveal that $\tau_L/T_0 \leq 2.5$ (where τ_L is the pulse width, full-width at half-maximum) is required to produce a soft-X-ray continuum of sizeable relative bandwidth ($>10\%$; that is, $>10 \text{ eV}$ in the 100-eV range). For pulses that meet this requirement and have a duration close to their bandwidth limit, the above-described dependence of the spectral structure of the cut-off radiation on φ is found to be robust within a broad range of parameters. Consequently, it can be used for calibrating the value of the C-E phase.

We have generated coherent soft X-rays by gently focusing our phase-stabilized 5-fs pulses into a 2-mm-long sample of neon gas. Figure 4 shows a series of soft-X-ray spectra produced under the conditions that are described in Fig. 4 legend for different values of the C-E phase of the 5-fs pump pulses. For a setting of $\varphi = \varphi_0$ (Fig. 4b), a broad structureless continuum appears in the cut-off region ($\hbar\omega > 120 \text{ eV}$). Notably, with a change of the phase, the continuous spectral distribution of the cut-off radiation gradually transforms into discrete harmonic peaks, with the maximum modulation depth appearing for the settings of $\varphi = \varphi_0 \pm \pi/2$. This behaviour is in agreement with the intuitive picture presented above, and allows us to identify φ_0 as zero with a residual ambiguity of $n\pi$, where n is an integer. This ambiguity in the determination of φ relates to the inversion symmetry of the interaction with the atomic gas medium. In fact, a π -shift in φ results in no change of the light waveform other than reversing the direction of the electromagnetic field vectors. This phase flip does not modify the intensity of the radiated X-ray photons, but it might become observable in photoelectron experiments^{8,10}.

Another noteworthy feature that has emerged from the soft-X-ray experiments with phase-stabilized pulses is the frequency shift of the harmonic peaks as a function of φ . This effect has been predicted by previous²⁷ simulations, as well as by our current simulations, and clearly manifests itself in the spectra shown in Fig. 4. This shift is due to a phase shift of the satellite burst with respect to the main X-ray burst, presumably as a consequence of the variation of the field amplitude within T_0 induced by that of the C-E phase. This behaviour reveals that the 'harmonic' fringes visible in the cut-off region of few-cycle-driven high-harmonic spectra do not represent genuine harmonics, because the latter should remain invariant to the change in the C-E phase as long as the laser frequency ω_L is left unchanged. This shift, induced by C-E phase, completely smears the harmonic structure of the near-cut-off soft X-rays generated by our 5-fs pulses in the absence of phase stabilization (Fig. 4e). Real high-order harmonics (invariant to shifts in φ) appear at photon energies some 40% below cut-off (plateau harmonics, not shown in Fig. 4), which are also observed in the absence of phase stabilization. The invariance of these plateau harmonics to shifts in φ confirms the fact that $\omega_L(t)$ is not affected by varying the C-E phase in our experiment.

The periodic spectral modulation present for $\varphi \approx \pi/2$ up to the highest photon energies observed, $\hbar\omega_{\text{cut-off}} \approx 130 \text{ eV}$ (Fig. 4d), completely disappears over a band of $\sim 16 \text{ eV}$ in the cut-off range as φ is shifted to zero (Fig. 4b). This is in fair agreement with the 20-eV continuum bandwidth predicted by our simple considerations for a 5-fs, 750-nm ($\tau_L/T_0 = 2$) laser pulse. The C-E phase can be evaluated with the highest possible accuracy by recording pairs of soft-X-ray harmonic spectra at $\varphi_1 = \varphi$ and $\varphi_2 = \varphi + \pi/2$, where φ is varied in small steps. In this method, $\varphi = 0$ (or, equivalently, π) can be identified from a pair of spectra that exhibit, respectively, the smallest and the largest modulation depth in the cut-off range. The current phase and intensity stability of our few-cycle pulses permits identification of a 'cosine' pulse with an accuracy better than $\pi/5$. The obtained value of φ is subsequently used to calibrate the interference pattern recorded in the second f-to-2f interferometer. With this calibration done, we can change φ with an accuracy of better than $\pi/10$ using the interferometer.

In this work we have demonstrated the generation of intense few-cycle light pulses with reproducible temporal evolution of the electromagnetic field. With these controlled intense light waveforms, we have explored the sensitivity of microscopic atomic currents to the timing of light field oscillations with respect to the pulse peak. The spectral distribution of the soft-X-ray radiation emitted by the atomic currents change as this timing is changed by less than 250 as, allowing a calibration of the C-E phase with a precision better than $\pi/5$. Intense few-cycle light pulses with a reproducible and known waveform allow precise control of the emission of an electron wave packet from an atom (or molecule) and its subsequent motion. The emerging field of attosecond physics is expected to benefit from this capability in several ways. First, the re-colliding electron wave packet may serve as a sub-femtosecond excitation pulse for atomic electron dynamics, or as a source of well-controlled, isolated sub-femtosecond X-ray pulses (Fig. 3e). Second, measurement of the shape and chirp^{28,29} of sub-femtosecond pulses is becoming feasible by recently demonstrated attosecond diagnostic techniques^{16,17}. Third, time-resolved inner-shell spectroscopy³⁰ could now be extended to tracing atomic processes at an attosecond timescale. Last, extending intense light waveform control to ultrahigh intensities would allow the motion of relativistic electrons to be steered with unprecedented accuracy. □

Methods

C-E phase stabilization loop

The high-power laser set-up (Fig. 2a) incorporates a phase-stabilized 10-fs laser system (FS 800, Menlo Systems). It consists of a dispersive-mirror-controlled Kerr-lens mode-locked Ti:sapphire oscillator (Femtosecond Compact Pro, Femtolasers), an f -to- $2f$ unit ('interferometer I' in Fig. 2a), and phase-locking electronics driving an acousto-optic modulator (AOM in Fig. 2a). About 50% of the output of the oscillator is directed into a 2-cm-long photonic crystal fibre (PCF in Fig. 2a) to broaden the spectrum of the pulses beyond an octave. Subsequently, the low-frequency components f_{low} of the broadened spectrum are frequency-doubled to produce an interference beat with the fundamental light at the high-frequency end of the spectrum, that is, $f_{\text{high}} = 2f_{\text{low}}$. The optical set-up producing this beat signal is referred to as an f -to- $2f$ interferometer. The phase of the two interfering quasi-monochromatic wave packets differs by $2\varphi - \varphi + \phi$, where ϕ is an unknown constant phase shift that prevents access to the value of φ in this type of experiment. For practical reasons, ϕ cannot be reliably calibrated from an f -to- $2f$ set-up. Despite the fact that the value of the C-E phase remains unknown, the information on its drift is sufficient to phase-stabilize the oscillator. In our laser system, the spectrally broadened oscillator pulses are fed into a Mach-Zehnder set-up (MZ in Fig. 2a) that has a frequency-doubling crystal in one of the arms^{2,20,21}. The interference with the original spectrum leads to the f -to- $2f$ beat signal $f_{\text{ce}} = \Delta\varphi f_{\text{rep}}/2\pi$ at around 530 nm, where $\Delta\varphi$ is a pulse-to-pulse shift of the C-E phase. The phase detector compares the electric signal detected at the output of the MZ with a reference, the frequency of which is derived by dividing the repetition rate of the oscillator ($f_{\text{rep}} \approx 80$ MHz) by a factor of 4. Forcing the two signals, $1/4 f_{\text{rep}}$ and ω_{CE} to oscillate in phase yields $\Delta\varphi = 1/4 \times 2\pi$ so that every fourth pulse looks alike. The rough adjustment of $\Delta\varphi$ is performed by changing the optical path length through a thin intracavity fused-silica plate. At this stage, φ is approximately reproduced in every fourth round trip in the laser cavity, which results in a quasi-periodic modulation of the f -to- $2f$ beat signal at $f_{\text{ce}} \approx f_{\text{rep}}/4 \approx 20$ MHz. Next, to enhance the accuracy of the phase reproducibility, the f -to- $2f$ beat signal is phase-locked to the $f_{\text{rep}}/4$ reference. This phase-locking is achieved by controlling $\Delta\varphi$ via nonlinear effects in the Ti:Sa crystal through variation of the pump intensity². For this purpose, the beam of the continuous-wave (c.w.) pump laser is passed through the AOM, which varies the power of the transmitted beam proportionately to the phase error signal described above. As a result, every fourth pulse, in the 80-MHz pulse train, and therefore also every 80,000th pulse picked by the multipass amplifier, carries the same C-E phase, giving rise to an accurate optical field reproduction in the radiation seeded into the amplifier. To correct additional C-E phase drift emerging during the process of amplification, the spectral interferograms detected behind the amplifier in the second f -to- $2f$ unit ('interferometer II' in Fig. 2a) are analysed in a personal computer by a Fourier-transform algorithm. The resultant (slowly varying) phase error signal is subsequently combined with the feedback from the phase detector in the first f -to- $2f$ interferometer and used to control the AOM.

Measurement of phase jitter introduced by the hollow waveguide

The method of (linear) spectral interference was used to assess the C-E phase jitter added by the hollow-fibre chirped-mirror pulse compressor (Fig. 2a). The measurement was performed by splitting off a reference beam behind the multipass amplifier, and interfering it with the spectrally broadened and temporally compressed light. The resultant spectral interferograms, recorded with a CCD (charge coupled device)-based spectrometer, reflected the shot-to-shot phase jitter that has two distinct contributions. The first arises from the mechanical path-length fluctuations of the interferometer arms, and corresponds to a linear variation of the spectral phase. The second represents the shift of the C-E phase resulting from the beam pointing and intensity instabilities of the light coupled into the

hollow waveguide. As the C-E phase jitter is frequency-independent, it can be clearly separated from the contribution added by the mechanical arm length drift.

Numerical simulations of the generation and propagation of soft X-rays

We used an improved version of the code first presented in refs 25 and 26. The code solves the wave propagation equations in three dimensions of space, where cylindrical symmetry is assumed for the transverse beam profile, and the slowly evolving wave approximation⁹ is used. The atomic dipole response to the laser field was calculated by an extension of the model of Lewenstein *et al.*^{14,25}, using accurate static field ionization rates obtained from numerical fully quantum-mechanical calculations. Harmonic absorption and diffraction is taken into account, as well as the geometrical (Gouy) phase shift, dispersion introduced by free electrons, ionization energy losses, and the delay due to the refractive index of neutral atoms.

Received 8 November 2002; accepted 3 January 2003; doi:10.1038/nature01414.

- Steinmeyer, G., Sutter, D., Gallmann, L., Matuschek, N. & Keller, U. Frontiers in ultrashort pulse generation: Pushing the limits in linear and nonlinear optics. *Science* **286**, 1507–1512 (1999).
- Xu, L., Spielmann, C., Poppe, A., Brabec, T. & Krausz, F. Route to phase control of ultrashort light pulses. *Opt. Lett.* **21**, 2008–2010 (1996).
- Krausz, F., Brabec, T., Schnürer, M. & Spielmann, C. Extreme nonlinear optics: Exploring matter to a few periods of light. *Opt. Photon. News* **9**, 46–51 (1998).
- de Bohan, A., Antoine, P., Milošević, D. B. & Piraux, B. Phase-dependent harmonic emission with ultrashort laser pulses. *Phys. Rev. Lett.* **81**, 1837–1840 (1998).
- Cormier, E. & Lambropoulos, P. Effect of the initial phase of the field in ionization by ultrashort laser pulses. *Eur. Phys. J. D* **2**, 15–20 (1998).
- Tempea, G., Geissler, M. & Brabec, T. Phase sensitivity of high-order harmonic generation with few-cycle laser pulses. *J. Opt. Soc. Am. B* **16**, 669–673 (1999).
- Christov, I. P. Phase-dependent loss due to nonadiabatic ionization by sub-10-fs pulses. *Opt. Lett.* **24**, 1425–1427 (1999).
- Dietrich, P., Krausz, F. & Corkum, P. B. Determining the absolute carrier phase of a few-cycle laser pulse. *Opt. Lett.* **25**, 16–18 (2000).
- Brabec, T. & Krausz, F. Intense few-cycle laser fields: Frontiers of nonlinear optics. *Rev. Mod. Phys.* **72**, 545–591 (2000).
- Paulus, G. G. *et al.* Absolute-phase phenomena in photoionization with few-cycle laser pulses. *Nature* **414**, 182–184 (2001).
- Corkum, P. B. Plasma perspective on strong-field multiphoton ionization. *Phys. Rev. Lett.* **71**, 1994–1997 (1993).
- Niikura, H. *et al.* Using correlated pairs for attosecond-resolution molecular wave packet measurements. *Nature* (in the press).
- Schäfer, K. J., Yang, B., DiMauro, L. F. & Kulander, K. C. Above threshold ionization beyond the high harmonic cutoff. *Phys. Rev. Lett.* **70**, 1599–1602 (1993).
- Lewenstein, M., Balci, P., Ivanov, M. Y., L'Huillier, A. & Corkum, P. B. Theory of high-harmonic generation by low-frequency laser fields. *Phys. Rev. A* **49**, 2117–2132 (1994).
- Paul, P. M. *et al.* Observation of a train of attosecond pulses from high harmonic generation. *Science* **292**, 1689–1692 (2001).
- Hentschel, M. *et al.* Attosecond metrology. *Nature* **414**, 509–513 (2001).
- Kienberger, R. *et al.* Steering attosecond electron wave packets with light. *Science* **297**, 1144–1148 (2002).
- Udem, T. *Phasenkohärente optische Frequenzmessungen am Wasserstoffatom* Thesis, Ludwig-Maximilians-Univ. (1997).
- Telle, H. R. *et al.* Carrier-envelope offset phase control: A novel concept for absolute optical frequency measurement and ultrashort pulse generation. *Appl. Phys. B* **69**, 327–332 (1999).
- Jones, D. J. *et al.* Carrier-envelope phase control of femtosecond mode-locked lasers and direct optical frequency synthesis. *Science* **288**, 635–639 (2000).
- Apolonski, A. *et al.* Controlling the phase evolution of few-cycle light pulses. *Phys. Rev. Lett.* **85**, 740–743 (2000).
- Sartania, S. *et al.* Generation of 0.1-TW 5-fs optical pulses at a 1-kHz repetition rate. *Opt. Lett.* **22**, 1562–1564 (1997).
- Kakehata, M. *et al.* Single-shot measurement of carrier-envelope phase changes by spectral interferometry. *Opt. Lett.* **26**, 1436–1438 (2001).
- Baltuška, A., Fujii, T. & Kobayashi, T. Self-referencing of the carrier-envelope slip in a 6-fs visible parametric amplifier. *Opt. Lett.* **27**, 1241–1243 (2002).
- Milosevic, N., Scrinzi, A. & Brabec, T. Numerical characterization of high harmonic attosecond pulses. *Phys. Rev. Lett.* **88**, 093905 (2002).
- Yakovlev, V. S. & Scrinzi, A. High harmonic imaging of few-cycle laser pulses. *Phys. Rev. Lett.* (submitted).
- Priori, E. *et al.* Nonadiabatic three-dimensional model of high-order harmonic generation in the few-optical-cycle regime. *Phys. Rev. A* **61**, 063801 (2000).
- Itatani, J. *et al.* Attosecond streak camera. *Phys. Rev. Lett.* **88**, 173903 (2002).
- Kitzler, M., Milosevic, N., Scrinzi, A., Krausz, F. & Brabec, T. Quantum theory of attosecond XUV pulse measurement by laser dressed photoionization. *Phys. Rev. Lett.* **88**, 173904 (2002).
- Drescher, M. *et al.* Time-resolved atomic inner-shell spectroscopy. *Nature* **419**, 803–807 (2002).

Acknowledgements We thank H. A. Haus, M. S. Pshenichnikov, G. F. Tempea and R. Kienberger for discussions, and Ch. Warmuth, Z. Cheng and M. Wieland for technical assistance. This work was sponsored by the Fonds zur Förderung der wissenschaftlichen Forschung (Austria) and the European ATTO network. We also thank Femtolasers GmbH and Menlo Systems GmbH for support.

Competing interests statement The authors declare competing financial interests: details accompany the paper on *Nature's* website (► <http://www.nature.com/nature>).

Correspondence and requests for materials should be addressed to F.K. (e-mail: ferenc.krausz@tuwien.ac.at).

A photovoltaic device structure based on internal electron emission

Eric W. McFarland* & Jing Tang†

* Department of Chemical Engineering, † Materials Department, University of California, Santa Barbara, California 93106-5080, USA

There has been an active search for cost-effective photovoltaic devices since the development of the first solar cells in the 1950s (refs 1–3). In conventional solid-state solar cells, electron–hole pairs are created by light absorption in a semiconductor, with charge separation and collection accomplished under the influence of electric fields within the semiconductor. Here we report a multilayer photovoltaic device structure in which photon absorption instead occurs in photoreceptors deposited on the surface of an ultrathin metal–semiconductor junction Schottky diode. Photoexcited electrons are transferred to the metal and travel ballistically to—and over—the Schottky barrier, so providing the photocurrent output. Low-energy (~ 1 eV) electrons have surprisingly long ballistic path lengths in noble metals^{4,5}, allowing a large fraction of the electrons to be collected. Unlike conventional cells, the semiconductor in this device serves only for majority charge transport and separation. Devices fabricated using a fluorescein photoreceptor on an Au/TiO₂/Ti multilayer structure had typical open-circuit photovoltages of 600–800 mV and short-circuit photocurrents of 10–18 $\mu\text{A cm}^{-2}$ under 100 mW cm^{-2} visible band illumination: the internal quantum efficiency (electrons measured per photon absorbed) was 10 per cent. This alternative approach to photovoltaic energy conversion might provide the basis for durable low-cost solar cells using a variety of materials.

The device structure is a solid-state multilayer; it consists of a photoreceptor layer deposited on a ~ 10 –50-nm Au film, which caps 200 nm of TiO₂ on an ohmic metal back contact (Fig. 1). Au and

TiO₂ are chosen and prepared so that a Schottky barrier, ϕ , of approximately 0.9 V is formed at the metal–semiconductor interface, and the dye, merbromin (2,7-dibromo-5-(hydroxymercurio)-fluorescein), is selected such that the photoexcited donor level is energetically above the barrier. The photon-to-electron conversion in this device occurs in four steps. First, light absorption occurs in the surface-absorbed photoreceptors, giving rise to energetic electrons. Second, electrons from the photoreceptor excited state are injected into conduction levels of the adjacent conductor, where they travel ballistically through the metal at an energy, ϵ_e , above the Fermi energy, E_f . Third, provided that ϵ_e is greater than the Schottky barrier height, ϕ , and the carrier mean-free path is long compared to the metal thickness, the electrons will traverse the metal and enter conduction levels of the semiconductor (internal electron emission). The absorbed photon energy is preserved in the remaining excess electron free energy when it is collected at the back ohmic contact, giving rise to the photovoltage, V . The photo-oxidized dye is reduced by transfer of thermalized electrons from states near E_f in the adjacent metal. Like electrochemical dye-sensitized solar cells^{6,7}, this structure physically separates the photon absorption process from charge separation and transport; but it does not have the disadvantage of needing a reducing agent in an electrolyte for intermolecular charge transport.

In addition to the desired production of ballistic electrons in the metal, there are other processes available for energy dissipation by the electron in the excited photoreceptor. The efficiency of the device will depend upon favouring specific energy transfer pathways from the major competing processes: (1) radiative intramolecular de-excitation with photon emission (luminescence), (2) non-radiative intramolecular or intermolecular de-excitation with direct coupling to phonons, (3) non-radiative intermolecular de-excitation with coupling to the metal conduction electrons, and (4) electron transfer of the energetic electron from the photoexcited state into the unoccupied conduction states of the metal. Both pathways (3) and (4) may produce energetic electrons in metal conduction levels above the Fermi energy (hot electrons) that also have sufficient energy and momentum to traverse the Schottky barrier; it is these electrons that give rise to the primary photocurrent in the device. We believe that the dominant pathway is (4) in our device, but we cannot rule out contributions to the measured photocurrent from pathway (3). Pathway (2) as well as non-transmitted electrons from (3) and (4) constitute the principal competing ‘quenching’ processes whereby the electronic excitation energy is dissipated as heat, and neither appears as light nor electron free energy in the device current.

The current–voltage characteristics of a representative device measured in the dark and under 100 mW cm^{-2} visible band

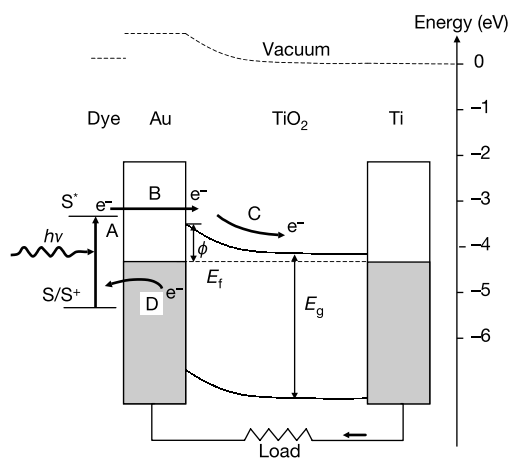


Figure 1 Electron transfer in the operating photovoltaic device. Process A, photon absorption and electron excitation from the chromophore ground state, S , to the excited state, S^* ; B, energetic electron transfer from S^* into and (ballistically) through the conducting surface layer and over the potential energy barrier into the semiconductor; C, electron conduction as a majority carrier within the semiconductor to the ohmic back contact and through the load; and D, reduction of the oxidized chromophore, S^+ , by a thermal electron from the conductor surface. Shown schematically are the relative energies of the electron levels within the device structures, the Schottky barrier, ϕ , the Fermi energy, E_f , and the semiconductor bandgap, E_g .

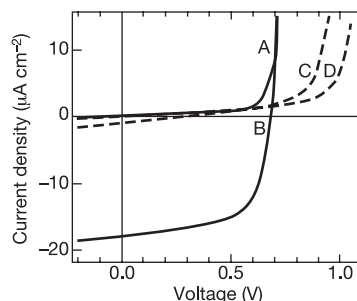


Figure 2 Current–voltage characteristics of the multilayer merbromin/Au/TiO₂/Ti photovoltaic device. The response is shown for dark conditions (curve A), and for 1,000 W m^{-2} broadband visible illumination (B). For comparison, curves obtained from the same device before adding merbromin are also shown for dark conditions (C), and under the same illumination (D).

illumination before and after applying merbromin are shown in Fig. 2. After deposition of merbromin, the surface-activated device had an open-circuit photovoltage, V_{oc} , of 685 mV, and a short-circuit photocurrent, j_{sc} , of $18.0 \mu\text{A cm}^{-2}$. The fill factor was determined to be 0.63 from the current–voltage characteristics under illumination. Before deposition of merbromin on the Au surface, excitation of defect levels in the thermally oxidized TiO_2 is responsible for a weak visible photoresponse ($V_{oc} = 325 \text{ mV}$, $j_{sc} = 0.9 \mu\text{A cm}^{-2}$).

The photovoltage, V , is the difference, under illumination, between the Fermi level of the semiconductor and the Fermi level of the ultrathin metal film. For the surface-sensitized Schottky diode, the expression for the maximum value under open circuit conditions, V_{oc} is identical to a conventional Schottky solar cell where photon absorption occurs by bandgap excitation:

$$V_{oc} = \left(\frac{nkT}{e} \right) [\ln(j_\gamma / (T^2 A^{**})) + e\phi_B / kT] \quad (1)$$

where n is the ideality factor (~ 1), k is Boltzmann's constant, e is the charge on the electron, A^{**} is the Richardson constant, and T is the device temperature⁵. The photocurrent density produced by photon absorption, j_γ , is balanced by the effective forward bias current from V_{oc} such that no net current is observed. The maximum photovoltages and photocurrents can be achieved by choice and preparation of the device materials to control and match the Schottky barrier height, ϕ , the semiconductor conduction band positions, and the chromophore donor/acceptor levels.

The photocurrent produced from a monochromatic photon flux, $F(\epsilon_\gamma)$, is determined largely by the photon capture efficiency and

the internal quantum efficiency, IQE, which is the number of hot electrons injected into the semiconductor (and thus detected) per absorbed photon:

$$j_\gamma(\epsilon_\gamma) = F(\epsilon_\gamma)\eta_\gamma(\epsilon_\gamma)\text{IQE}(\epsilon_\gamma) \quad (2)$$

$$\text{IQE}(\epsilon_\gamma) = \int_0^{\epsilon_\gamma} \eta_{CM}(\epsilon_\gamma, \epsilon_e)\eta_{MS}(\epsilon_e)\eta_S d\epsilon_e \quad (3)$$

where $\eta_\gamma(\epsilon_\gamma)$ is the efficiency for absorption of a photon of energy ϵ_γ ; $\eta_{CM}(\epsilon_\gamma, \epsilon_e)$ is the probability that absorption will result in the injection of an excited electron into the metal at an energy, ϵ_e , above the Fermi energy; $\eta_{MS}(\epsilon_e)$ is the efficiency for charge transport across the metal film and into the semiconductor; and η_S is the charge collection efficiency in the semiconductor. Experimentally, the incident photon-to-electron conversion efficiency, $\text{IPCE}(\epsilon_\gamma) = j(\epsilon_\gamma)/F(\epsilon_\gamma)$, is typically determined by measuring the photocurrent from a monochromatic photon source. The IQE is then calculated by correcting the IPCE for photon absorption in the dye as $\text{IQE}(\epsilon_\gamma) = \text{IPCE}(\epsilon_\gamma)/\eta_\gamma(\epsilon_\gamma)$.

What makes the IQE for this device structure acceptable is the surprisingly high value of $\eta_{MS}(\epsilon_e)$. At energies of $\sim 1 \text{ eV}$ above the Fermi level, the ballistic mean free path for electrons in Au and other metals with low-lying or filled d bands has been measured to be extremely long, $\sim 20\text{--}150 \text{ nm}$ (refs 4, 5). Thus, despite the competing processes of quasi-elastic scattering by phonons and inelastic scattering from other electrons, provided the metal conducting film is ultrathin, a significant fraction of the injected electrons will reach the Schottky barrier.

Although the relationship between V_{oc} and j_γ in our devices is identical to that of a conventional semiconductor Schottky solar cell, equation (1), the bulk semiconductor is not utilized for photon absorption; thus the bandgap and semiconductor thickness constraints are largely removed. The significance of the Schottky barrier height, however, is increased. The absorption process occurs in the photoreceptor, which is selected to balance high solar-spectrum absorbance with maximum photovoltage as well as chemical stability. In this design, several different chromophores could be used simultaneously to provide more efficient conversion (for example, dyes, quantum structures); however, in an idealized cell using a single dye, the optimal chromophore absorption maximum would be the same as the ideal bandgap of a conventional Schottky diode solar cell ($\sim 1.5 \text{ eV}$; refs 2, 8). Thus, the theoretical maximum power output of the Schottky device utilizing internal electron emission can be no greater than an idealized conventional cell for the solar spectrum, $\sim 25\%$ (refs 2, 8).

The physical and electronic coupling of the chromophore to the metal conduction levels are crucial to the performance of the device. In aqueous solution, merbromin has an absorption maximum at 511 nm wavelength; but when merbromin is attached to the Au film, the primary absorption peak appears to split into a blue-shifted peak and a stronger red-shifted peak, giving rise to a broadening and red-shift of the overall spectrum (Fig. 3a). The wavelength-dependent photoresponse of the device is shown in Fig. 3b, where we plot the IPCE under short-circuit conditions, $\text{IPCE} = j(\epsilon_\gamma)/F(\epsilon_\gamma)$, equation (2). The general features of the IPCE response spectrum, and the broad maximum, are approximately the same as the surface-bound dye absorption; this is consistent with the mechanism of action described in Fig. 1, with both the associated red- and blue-shifted dye states coupling to the metal such that hot electrons are produced. The IQE is determined by correcting the IPCE for the fraction of incident photons absorbed in the dye (Fig. 3c). Although the broad absorption edge of the TiO_2 overlaps with the dye adsorption between 400 nm and 450 nm, above 500 nm the TiO_2 absorption is negligible and the IQE of approximately 10% reflects the efficiency of the internal emission process.

The main considerations in choosing the materials for the device are: (1) the relative energies of the donor/acceptor levels in the

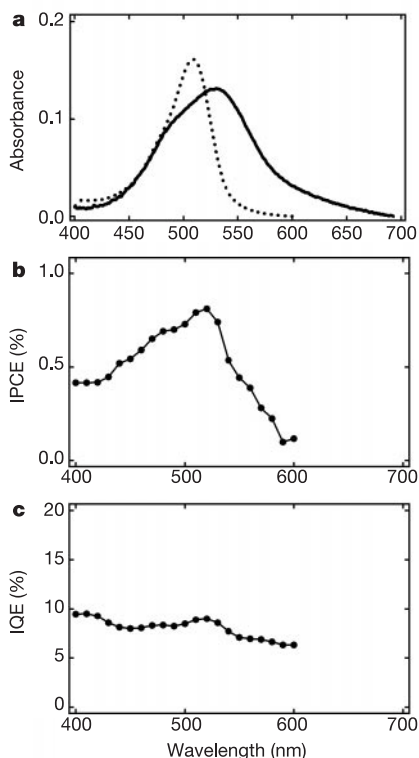


Figure 3 The wavelength-dependent photoresponse and absorbance of the photovoltaic device. **a**, The absorbance of merbromin adsorbed to the surface of the Au/TiO₂/Ti device (solid line). Shown for reference is the absorbance (in arbitrary units) of merbromin in water (dotted line). **b**, The incident photon-to-electron conversion efficiency (IPCE) of the dye-sensitized device under 0.4 mW cm^{-2} illumination. **c**, The internal quantum efficiency (IQE), determined by correcting the IPCE by the absorption efficiency of the affixed dye, reflects the absorbed photon-to-electron conversion efficiency of the dye.

photoreceptor, the conductor Fermi level, the barrier height, and the position of the semiconductor conduction band edges, (2) the ballistic mean free path of electrons, and (3) the physical and electronic coupling of the chromophore to the conductor for high absorbance and high electron transfer efficiency. The devices studied here represent only one of several different configurations of photovoltaics taking advantage of ballistic electron transport and internal electron emission in a photovoltaic device. For example, modifications of the structure to utilize hot hole injection (rather than hot electrons) in a p-type junction⁹ would allow use of hole conducting polymers instead of inorganic semiconductors.

The IPCE and overall energy efficiency are limited by low dye coverage (8×10^{14} molecules cm^{-2}) and the resulting low photon absorption. Significant increases are expected with improved optical design (reduced surface reflection), decreased metal thickness, increased dye loading, and an engineered surface morphology with significantly higher surface area structured such that multiple passes through a dye-covered surface are possible for each photon. Although the ultimate efficiency of an optimized device based on the concept presented here is approximately the same as an ideal conventional semiconductor cell, there appear to be practical and economic advantages in terms of the wide choice of inexpensive, durable, and readily synthesized device materials that may be utilized. □

Methods

Device fabrication

Devices were fabricated on titanium foil substrates (Alfa Aesar), which served as ohmic back contacts. A 250-nm layer of titanium (99.9999%) was evaporated under vacuum onto the foil following cleaning and polishing (using 10 μm grit). A 200-nm layer of TiO_2 was grown on the substrate by thermal oxidation at 500 °C. The polycrystalline TiO_2 is predominately rutile phase, with oxygen vacancies giving rise to n-type doping. Au films were electrodeposited onto the TiO_2 from a solution containing 0.2 M KCN and 0.1 M AuCN at pH 14. The TiO_2 served as the working electrode, with a Pt wire counter electrode. A 100-ms galvanostatic pulse at -200 mA cm^{-2} was used to nucleate Au uniformly on the surface, followed by a periodic galvanostatic pulse train of 5 ms at $+0.2 \text{ mA cm}^{-2}$ and 5 ms at -1.7 mA cm^{-2} for 10 s to form a film ~ 10 –50 nm thick. Photoactive merbromin (2,7-dibromo-5-(hydroxymethyl)fluorescein disodium salt, 5 mM in water) was adsorbed onto the surface by immersion at room temperature for 10–12 h, followed by rinsing in water.

Characterization

Current–voltage (I – V) curves were measured using a voltage ramp rate of 0.05 V s^{-1} in the dark and under illumination from a 250-W tungsten lamp (Oriel, 6129), with intensity measured using a radiometer (IL1700, International Light). The fill factor was calculated at $1,000 \text{ W m}^{-2}$ by dividing the maximum product of current and voltage from the illuminated I – V curve by the product of open-circuit voltage and short-circuit current at the same illumination. The spectral response was determined using a 150-W Xe lamp and monochromator (Oriel 7240). IPCE was calculated from the current density under short-circuit conditions and the photon flux as measured by the radiometer. The optical absorbance (and absorption efficiency, $\eta_{\text{ph}}(\epsilon_{\text{ph}})$) of the dye on the device surface and dye photon absorption was determined from the transmission and reflectance of a device fabricated on a transparent substrate before and after application of the dye, using an integrating sphere (LabSphere) and fibre-optic coupled monochromator (Ocean Optics). Free-solution dye absorbance was measured with an optical spectrometer (UV-1610, Shimadzu). Dye loading was determined by detaching the dye from the activated device surface in 1 mM NaOH solution, and determining the amount removed from the difference in optical absorbance at 511 nm of the NaOH solution.

Received 17 July; accepted 18 November 2002; doi:10.1038/nature01316.

- Chapin, D. M., Fuller, C. S. & Pearson, G. L. A new silicon p-n junction photocell for converting solar radiation into electrical power. *J. Appl. Phys.* **25**, 676–677 (1954).
- Archer, M. D. & Hill, R. (eds) *Clean Electricity from Photovoltaics* (Series on Photoconversion of Solar Energy, Vol. 1, Imperial College Press, London, 2001).
- Goetzberger, A. & Hebling, C. Photovoltaic materials, past, present, future. *Sol. Energy Mater. Sol. Cells* **62**, 1–19 (2000).
- Seah, M. P. & Dench, W. A. Quantitative electron spectroscopy of surfaces: a standard data base for electron inelastic mean free paths in solids. *Surf. Interf. Anal.* **1**, 2–11 (1979).
- Frese, K. W. & Chen, C. Theoretical models of hot carrier effects at metal-semiconductor electrodes. *J. Electrochem. Soc.* **139**, 3234–3249 (1992).
- Grätzel, M. Photoelectrochemical cells. *Nature* **414**, 338–344 (2001).
- Green, M. A., Emery, K., King, D. L., Igari, S. & Warta, W. Solar cell efficiency tables (version 17). *Prog. Photovolt. Res. Appl.* **9**, 49–56 (2001).
- Se, S. M. *Physics of Semiconductor Devices* 2nd edn, Ch. 14 (Wiley & Sons, New York, 1981).
- Nienhaus, H. et al. Electron-hole pair creation at Ag and Cu surfaces by adsorption of atomic hydrogen and deuterium. *Phys. Rev. Lett.* **82**, 446–448 (1999).

Acknowledgements We thank M. White, A. Tavakkoly, A. Kochhar, N. Shigeoka, G. Stucky and W. Siripala for technical assistance and discussions. The project was supported by Adrena Inc. Financial support for J.T. was provided by the NSF-MRSEC funded Materials Research Laboratory (UCSB).

Competing interests statement The authors declare competing financial interests: details accompany the paper on *Nature's* website (► <http://www.nature.com/nature>).

Correspondence and requests for materials should be addressed to E.W.M. (e-mail: mcfar@engineering.ucsb.edu).

Hydrothermal recharge and discharge across 50 km guided by seamounts on a young ridge flank

A. T. Fisher^{*†}, E. E. Davis[‡], M. Hutnak^{*}, V. Spiess[§], L. Zühlsdorff[§], A. Cherkaoui^{*}, L. Christiansen^{||}, K. Edwards[¶], R. Macdonald[‡], H. Villinger[§], M. J. Mottl[#], C. G. Wheat[☆] & K. Becker[¶]

^{*} Earth Sciences Department, [†] Institute for Geophysics and Planetary Physics, University of California, Santa Cruz, California 95064, USA

[‡] Pacific Geoscience Center, Geological Survey of Canada, Sydney, British Columbia V8L 4B2, Canada

[§] Earth Sciences Department, University of Bremen, Bremen D-28359, Germany

^{||} Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, Maryland 21218, USA

[¶] Rosenstiel School of Marine and Atmospheric Science, University of Miami, Miami, Florida 33149, USA

[#] School of Earth and Ocean Science and Technology, University of Hawaii, Honolulu, Hawaii 96822, USA

[☆] Global Undersea Research Unit, University of Alaska, Fairbanks, Alaska 99775, USA

Hydrothermal circulation within the sea floor, through lithosphere older than one million years (Myr), is responsible for 30% of the energy released from plate cooling, and for 70% of the global heat flow anomaly (the difference between observed thermal output and that predicted by conductive cooling models)^{1,2}. Hydrothermal fluids remove significant amounts of heat from the oceanic lithosphere for plates typically up to about 65 Myr old^{3,4}. But in view of the relatively impermeable sediments that cover most ridge flanks⁵, it has been difficult to explain how these fluids transport heat from the crust to the ocean. Here we present results of swath mapping, heat flow, geochemistry and seismic surveys from the young eastern flank of the Juan de Fuca ridge, which show that isolated basement outcrops penetrating through thick sediments guide hydrothermal discharge and recharge between sites separated by more than 50 km. Our analyses reveal distinct thermal patterns at the sea floor adjacent to recharging and discharging outcrops. We find that such a circulation through basement outcrops can be sustained in a setting of pressure differences and crustal properties as reported in independent observations and modelling studies.

Hydrothermal circulation on ridge flanks (crust older than 1 Myr) advects lithospheric heat from much of the sea floor, contributing to enormous fluxes of fluid, energy and solutes^{1,6,7}. It is easy for fluid to enter and leave the crustal reservoir on most young sea floor, where sediment cover is incomplete and permeable basement rocks are widely exposed, but mechanisms by which fluids penetrate through thick and more continuous sediments have remained enigmatic⁵. The primary difficulty is that forces available to drive hydrothermal circulation on ridge flanks are modest^{7–10}, being limited mainly to the difference in fluid pressures below

columns of recharging (cool) and discharging (warm) fluid. Sediment permeabilities are typically so low that even a few tens of metres of sediment cover will reduce seepage to rates that are incapable of liberating significant quantities of lithospheric heat^{7,8}. Even sites located below areas of relatively rapid sedimentation may lose a considerable fraction of their heat advectively¹¹, but it has not been shown previously how hydrothermal fluids move across the thick sediment layer at rates sufficiently rapid to efficiently cool the oceanic crust over vast areas.

Ocean Drilling Program (ODP) Leg 168 investigated ridge-flank hydrothermal processes along an 80-km transect of sites east of the Juan de Fuca ridge¹² (Fig. 1a). Sediment cover over basement in this area is almost continuous at distances >20 km from the active spreading centre because of proximity to North America, a source for turbidites during Pleistocene glacial maxima. Pore fluids collected from just above basement along this transect¹², and recovered from an overpressured hole at Site 1026 that penetrated a buried basement ridge below 250 m of sediment¹³ (Fig. 1b), were analysed for composition and ¹⁴C age¹⁴. There is a general progression in fluid evolution along the western part of the drilling transect, as

reaction temperatures rise, and ¹⁴C ages increase from 1.0 to 9.9 thousand years (kyr) (relative to present bottom water) at distances of 3.3 to 14.6 km from the point of sediment onlap, respectively (Fig. 1a). But fluids collected from Site 1026 at the eastern end of the transect have a ¹⁴C age of only 4.3 kyr, indicating that these waters could not have recharged near the western end of the drilling transect.

Swath map coverage of the region is extensive (Fig. 1b), and there are no known outcrops between exposed basement close to the spreading centre and Site 1026 that could have permitted recharge of younger water. There are several basaltic outcrops near Site 1026 (Fig. 1b) that are known sites of hydrothermal discharge^{15–17}. The Baby Bare outcrop has been surveyed by surface ship, submersible, and remotely operated vehicle, and numerous vent sites have been located^{16,18}. Baby Bare heat output is 2–20 MW, and the discharge flux of hydrothermal fluid is 4–13 l s⁻¹ (refs 16, 19). Although Baby Bare rises just 65 m above the surrounding sea floor (Fig. 1c), the basaltic edifice is substantial, rising 600 m above regional basement (Fig. 2a).

Co-located heat-flow and seismic measurements across Baby

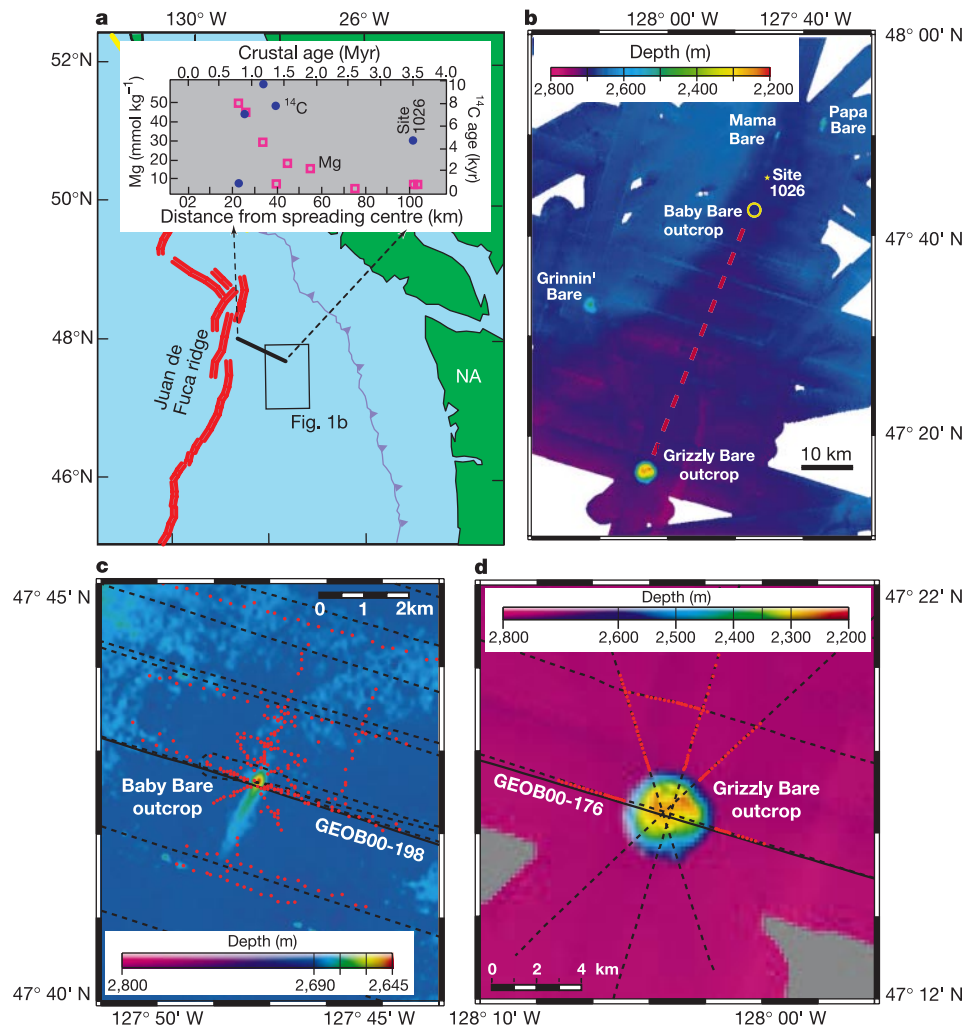


Figure 1 Maps of field area on the eastern flank of Juan de Fuca ridge, and selected geochemical data collected during ODP Leg 168¹². **a**, Regional index map showing spreading ridges (red) and subduction zone (purple). NA, North America. Thick black line is Leg 168 transect. Inset, Mg concentrations and ¹⁴C ages from upper basement pore fluids¹⁴. **b**, Swath map bathymetry showing locations of outcrops and ODP Site 1026 (star). Dashed red line indicates buried basement ridge, subparallel to spreading centre to

the west, that runs between several outcrops and ODP Site 1026. **c**, Bathymetric detail around Baby Bare outcrop, showing locations of seismic profiles (black lines) and heat-flow measurements (red dots). Seismic profile GEOB00-198 (solid line) is shown in Fig. 2a. **d**, Bathymetric detail around Grizzly Bare outcrop. Symbols as in **c**. Seismic profile GEOB00-176 is shown in Fig. 2b.

Bare illustrate the local influence of fluid venting on the thermal state of the crust (Fig. 2a). Heat flow away from the outcrop is consistent with conductive cooling models for sea floor of this age, after correcting for sedimentation²⁰. Heat flow is considerably greater within a few hundred metres of the outcrop, rising above 1 W m^{-2} . We have calculated isotherm depths near the outcrop by continuing sea-floor heat-flow values downwards to upper basement, on the basis of thermal and seismic data (Fig. 2a). Isotherms are generally subparallel to the sediment–basement interface until very close to the outcrop, with an upper basement temperature of about 65°C , near that measured at Site 1026¹³. The rise in heat flow near the outcrop results from maintenance of nearly isothermal basement temperatures as sediment thins, indicating extremely vigorous local circulation.

Fluid seepage downwards through sediments near Baby Bare outcrop cannot recharge warm springs discharging on this feature. The youthful ^{14}C age of shallow basement fluids¹⁴ would require a recharge velocity of at least 10 cm yr^{-1} , given typical sediment thickness. Geochemical analyses of pore fluids collected well away from basement highs in this area indicate diffusive and reactive conditions throughout the thick sediment layer, with no evidence for vertical advection^{12,21}. These analyses are sensitive to fluid velocities $\geq 0.1 \text{ mm yr}^{-1}$, and thus an area of $1,500 \text{ km}^2$ around Baby Bare outcrop (a radial distance of 22 km) would be required to support discharge of 5 l s^{-1} . Even assuming a recharge velocity of 0.1 mm yr^{-1} , sediment properties^{8,12} would require a much greater differential pressure than observed regionally^{8,9,22} to draw bottom sea water downward through the sediment layer and into basement.

Sulphate compositions of Baby Bare vent fluids require recharge through basaltic outcrops²¹. There is no thermal or geochemical evidence for hydrothermal recharge through outcrops close to Baby Bare, but a consistent trend in the composition of basement pore

fluids from south to north along the buried basement ridge below Baby Bare and Site 1026²³ indicates that recharge may occur to the south. We investigated the hydrogeology of the two closest outcrops south of Baby Bare. Heat-flow observations around the Grinnin' Bare outcrop, 35 km southwest of Baby Bare (Fig. 1b), are also indicative of fluid venting, but the Grizzly Bare outcrop, 52 km south-southwest of Baby Bare, hosts a contrasting thermal regime

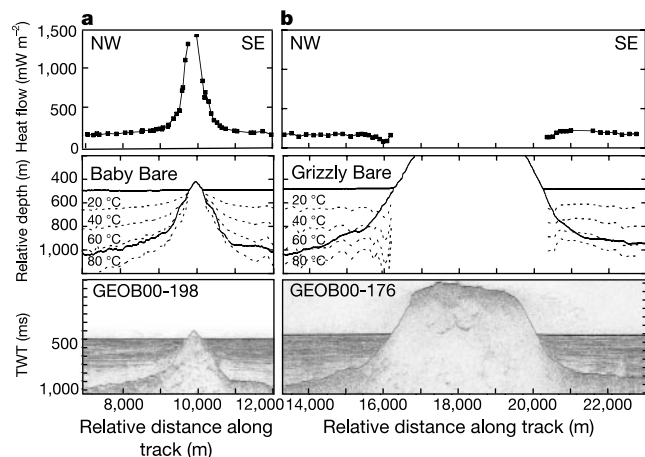


Figure 2 Heat-flow values, isotherms and seismic profiles across the Baby Bare and Grizzly Bare outcrops. Heat-flow values are those that are not on the outcrops and are within 100 m of the seismic lines. Typical sediment thickness away from the outcrops is about 500 m . Heat-flow values away from the outcrops are consistent with conductive cooling models for 3.5-Myr -old sea floor, after correcting for sedimentation²⁰, and upper basement temperatures are about 65°C . **a**, Isotherms are subparallel to the sediment–basement interface in the vicinity of the Baby Bare outcrop, rising with basement abruptly near the outcrop. Isotherm locations are approximate, being based on the assumption of locally conductive conditions. **b**, Heat flow decreases near the edge of Grizzly Bare outcrop, and basement temperatures remain depressed out to a distance of several kilometres from the area of basement exposure. Local variations in heat flow—for example, the single elevated value adjacent to the northwest side of Grizzly Bare outcrop—are likely to result from irregularities in aquifer geometry and properties, and are most common where sediment cover is thin.

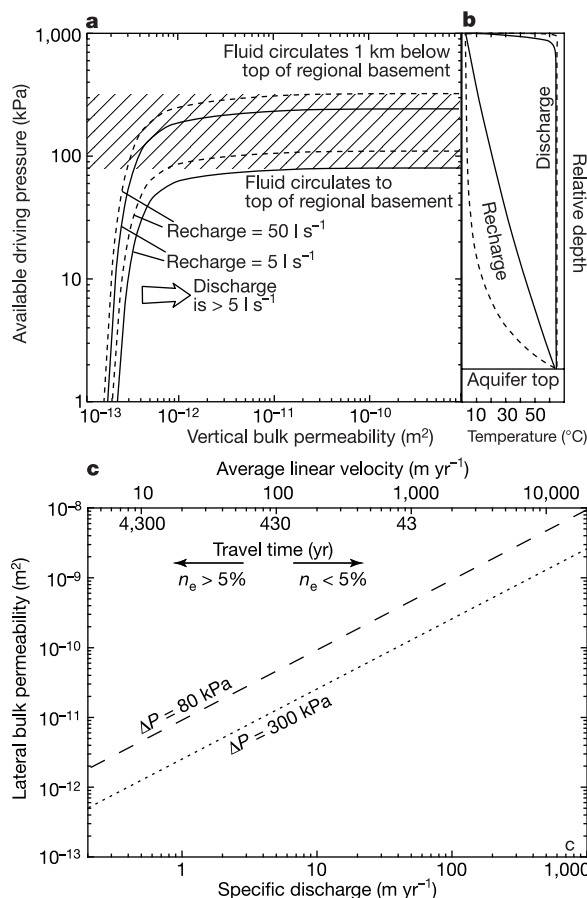


Figure 3 Calculated pressures available to drive large-scale lateral fluid flow, thermal profiles within outcrops, and lateral bulk permeabilities within basement between outcrops. **a**, Pressures available to drive large-scale lateral fluid flow within basement between Grizzly Bare (recharge) and Baby Bare (discharge) outcrops, based on the difference between pressures at the base of recharging and discharging columns of fluid. Calculations were completed for discharge of 5 l s^{-1} , and for recharge of 5 l s^{-1} (solid line) and 50 l s^{-1} (dashed line). The lower values in each case indicate results for fluid circulation to the top of regional basement, and the higher values show results for flow at a depth of 1 km below top of regional basement. Vertical permeabilities $< 10^{-12} \text{ m}^2$ would result in considerable energy being lost during ascent and descent, leaving less differential pressure to drive lateral flow at depth. Shaded band indicates range of driving pressures based on vertical permeability $> 10^{-12} \text{ m}^2$. Higher discharge values¹⁹ would move the curves to the right, as indicated by the arrow, but would not change the pressure available to drive large-scale, lateral flow. **b**, Vertical thermal profiles at recharge and discharge sites were calculated using a one-dimensional model of heat and fluid flow²⁴, and volume fluxes of 5 l s^{-1} (solid line) or 50 l s^{-1} (dashed line). The calculated discharge profiles are more isothermal than the recharge profiles because the fluid is assumed to pass through a much smaller cross-sectional area at Baby Bare outcrop than at Grizzly Bare outcrop. **c**, Lateral bulk permeability required between the outcrops as a function of volume flow/area. Travel times and average linear velocities shown on upper axis are for an effective porosity, n_e , of 5% ; higher or lower values would shift the upper axis as indicated with the arrows. The actual fluid travel time between outcrops cannot be greater than that the ^{14}C age of 4.3 kyr (ref. 14), but considerably shorter times are indicated once dispersive loss during flow is considered^{25–27}.

(Figs 1 and 2b). Heat flow away from Grizzly Bare is consistent with crustal age, but values drop abruptly by 50% adjacent to the outcrop. Isotherms within uppermost basement around Grizzly Bare are suppressed relative to those at Baby Bare out to several kilometres from the edge of basement exposure (Fig. 2b). These observations indicate that recharge of cold sea water cools the basement rocks. On the basis of geochemical observations^{14,23} and a lack of other recharge or discharge sites between the two outcrops, we infer that some of the water recharging Grizzly Bare flows north-northeast and vents at the Baby Bare outcrop 52 km away.

To test whether this hydrogeologic interpretation is reasonable, we quantify available forces and basement properties necessary to drive fluid flow at rates consistent with observations (Fig. 3). In the absence of recent volcanism¹⁸, driving forces responsible for large-scale lateral flow on ridge flanks are limited to the difference in pressure between recharging (cool) and discharging (warm) columns of water⁹, $\Delta P = g \int_0^z \Delta \rho(z) dz$, where g is the acceleration due to gravity, $\Delta \rho(z)$ is the difference in fluid density, and z is the height of recharging and discharging water columns. We calculate the pressure difference using a one-dimensional model of coupled heat and fluid flow with fixed upper and lower boundary temperatures²⁴, 2 °C and 65 °C, respectively. We use a hydrothermal discharge flux of 51 s^{-1} (ref. 16), and recharge fluxes of 51 s^{-1} and 501 s^{-1} . The first case presumes that all recharge at Grizzly Bare circulates to Baby Bare, whereas the second case supposes that only 10% of recharge follows that path (the rest venting elsewhere). We subtract the energy lost during vertical flow, using Darcy's law and assuming that recharge and discharge are distributed evenly across the outcrop surfaces. We make these calculations for various depths into basement, and show results for two extreme cases: fluid flows laterally at the top of regional basement (600 m below the sea floor), and fluid flows laterally 1,000 m into regional basement (1,600 m below the sea floor).

If vertical permeability within the two outcrops were $\leq 10^{-13} \text{ m}^2$, all energy available to drive lateral flow would be lost during fluid descent and ascent (Fig. 3a). At vertical permeabilities $\geq 10^{-12} \text{ m}^2$, little energy is lost during fluid descent and ascent, and 80–300 kPa could be available to drive lateral flow between the outcrops (Fig. 3a). The ^{14}C age of basement water from Site 1026 (4.3 kyr; ref. 14), which is geochemically very similar to Baby Bare vent fluids²³, provides an upper limit for the travel time between Grizzly Bare and Baby Bare. Studies of groundwater age demonstrate that large corrections are needed to account for dispersive losses between layers hosting the most significant flow and the stagnant layers around them^{25–27}. Particularly within heterogeneous, fractured aquifers, corrections of the order of 10 to 100 times or more are required. If the actual travel time between outcrops is 4.3 kyr (no dispersive loss), and the driving pressure is 80–300 kPa, lateral bulk permeability of $\geq 10^{-12} \text{ m}^2$ is required (Fig. 3c). If the dispersive correction is 100 times, then bulk permeability would need to be 10^{-10} m^2 and the travel time between outcrops would be only about 40 yr (Fig. 3c).

Permeability values in upper basement of $\geq 10^{-12} \text{ m}^2$ were determined by borehole measurements at Site 1026¹³, and values $\geq 10^{-10} \text{ m}^2$ were estimated on the basis of analysis of formation tidal response and considerations of thermal homogeneity^{10,22}. Thus there is sufficient basement permeability and a large enough driving force to allow water recharging at the Grizzly Bare outcrop to discharge at the Baby Bare outcrop, 52 km to the north-northwest. Bulk permeability may be enhanced in this direction (subparallel to the spreading axis to the west) by faults and fractures in the crust developed during and after crustal creation. Sea-floor heat-flow values away from outcrops in the study area do not indicate a significant regional heat-flow deficit, as observed on most young ridge flanks^{3,4}, so the fluid flux through upper basement must be relatively modest. This is probably a result of the thick and mostly continuous sediment cover that isolates much of the basement

aquifer from easy communication with the overlying ocean.

Seamounts and other basement outcrops are common features on a global basis^{28,29}, and the processes documented by this study illustrate how hydrogeologic communication between the hydrothermal entry and exit points is possible over large distances. This explains how much of the sea floor, which is typified by a greater density of basement outcrops than found on the eastern Juan de Fuca ridge flank, can experience pervasive advective heat loss even while sealed locally below thick sediments. □

Received 4 September; accepted 26 November 2002; doi:10.1038/nature01352.

1. Elderfield, H. & Schultz, A. Mid-ocean ridge hydrothermal fluxes and the chemical composition of the ocean. *Annu. Rev. Earth Planet. Sci.* **24**, 191–224 (1996).
2. Sclater, J. G., Jaupart, C. & Galson, D. The heat flow through oceanic and continental crust and the heat loss of the earth. *Rev. Geophys. Space Phys.* **18**, 269–311 (1980).
3. Parsons, B. & Sclater, J. G. An analysis of the variation of ocean floor bathymetry and heat flow with age. *J. Geophys. Res.* **82**, 803–829 (1977).
4. Stein, C. & Stein, S. A model for the global variation in oceanic depth and heat flow with lithospheric age. *Nature* **359**, 123–129 (1992).
5. Jacobson, R. S. Impact of crustal evolution on changes of the seismic properties of the uppermost oceanic crust. *Rev. Geophys.* **30**, 23–42 (1992).
6. Williams, D. L. & Von Herzen, R. P. Heat loss from the earth: new estimate. *Geology* **2**, 327–328 (1974).
7. Mottl, M. J. & Wheat, C. G. Hydrothermal circulation through mid-ocean ridge flanks: fluxes of heat and magnesium. *Geochim. Cosmochim. Acta* **58**, 2225–2237 (1994).
8. Giambalvo, E., Fisher, A. T., Darty, L., Martin, J. T. & Lowell, R. P. Origin of elevated sediment permeability in a hydrothermal seepage zone, eastern flank of Juan de Fuca Ridge, and implications for transport of fluid and heat. *J. Geophys. Res.* **105**, 913–928 (2000).
9. Fisher, A. & Becker, K. Channelized fluid flow in oceanic crust reconciles heat-flow and permeability data. *Nature* **403**, 71–74 (2000).
10. Davis, E. E. & Becker, K. Observations of natural-state fluid pressures and temperatures in young oceanic crust and inferences regarding hydrothermal circulation. *Earth. Planet. Sci. Lett.* **204**, 231–248 (2002).
11. Baker, P., Stout, P., Kastner, M. & Elderfield, H. Large-scale lateral advection of seawater through oceanic crust in the central equatorial Pacific. *Earth. Planet. Sci. Lett.* **105**, 522–533 (1991).
12. Davis, E. E., Fisher, A. T., Firth, J. *Proc. ODP Init. Rep. Leg* **168** (1997).
13. Fisher, A. T., Becker, K. & Davis, E. E. The permeability of young oceanic crust east of Juan de Fuca Ridge determined using borehole thermal measurements. *Geophys. Res. Lett.* **24**, 1311–1314 (1997).
14. Elderfield, H., Wheat, C. G., Mottl, M. J., Monnin, C. & Spiro, B. Fluid and geochemical transport through oceanic crust: a transect across the eastern flank of the Juan de Fuca Ridge. *Earth. Planet. Sci. Lett.* **172**, 151–165 (1999).
15. Davis, E. E. *et al.* FlankFlux: an experiment to study the nature of hydrothermal circulation in young oceanic crust. *Can. J. Earth Sci.* **29**, 925–952 (1992).
16. Mottl, M. J. *et al.* Warm springs discovered on 3.5 Ma oceanic crust, eastern flank of the Juan de Fuca Ridge. *Geology* **26**, 51–54 (1998).
17. Wheat, C. G. & Mottl, M. Composition of pore and spring waters from Baby Bare: global implications of geochemical fluxes from a ridge flank hydrothermal system. *Geochim. Cosmochim. Acta* **64**, 629–642 (2000).
18. Becker, N. C., Wheat, C. G., Mottl, M. J., Karsten, J. & Davis, E. E. A geological and geophysical investigation of Baby Bare, locus of a ridge flank hydrothermal system in the Cascadia Basin. *J. Geophys. Res.* **105**, 23557–23568 (2000).
19. Thomson, R. E., Davis, E. E. & Burd, B. J. Hydrothermal venting and geothermal heating in Cascadia Basin. *J. Geophys. Res.* **100**, 6121–6141 (1995).
20. Davis, E. E. *et al.* Regional heat-flow variations across the sedimented Juan de Fuca Ridge eastern flank: constraints on lithospheric cooling and lateral hydrothermal heat transport. *J. Geophys. Res.* **104**, 17675–17688 (1999).
21. Wheat, C. G. & Mottl, M. J. Hydrothermal circulation, Juan de Fuca Ridge eastern flank: factors controlling basement water composition. *J. Geophys. Res.* **99**, 3067–3080 (1994).
22. Davis, E. E., Wang, K., Becker, K. & Thompson, R. E. Formation-scale hydraulic and mechanical properties of oceanic crust inferred from pore-pressure response to periodic seafloor loading. *J. Geophys. Res.* **105**, 13423–13435 (2000).
23. Wheat, C. G., Elderfield, H., Mottl, M. J. & Monnin, C. Chemical composition of basement fluids within an oceanic ridge flank: implications for along-strike and across-strike hydrothermal circulation. *J. Geophys. Res.* **105**, 13437–13447 (2000).
24. Bredehoeft, J. D. & Papadopoulos, I. S. Rates of vertical groundwater movement estimated from the earth's thermal profile. *Wat. Resour. Res.* **1**, 325–328 (1965).
25. Sanford, W. E. Correcting for diffusion in carbon-14 dating of ground water. *Ground Wat.* **35**, 357–361 (1997).
26. Bethke, C. M. & Johnson, T. M. Paradox of groundwater age. *Geology* **30**, 107–110 (2002).
27. Shapiro, A. M. Effective matrix diffusion in kilometer-scale transport in fractured rock. *Wat. Resour. Res.* **37**, 507–522 (2001).
28. Wessel, P. Global distribution of seamounts inferred from gridded Geosat/ERS-1 altimetry. *J. Geophys. Res.* **106**, 19431–19442 (2001).
29. Villinger, H., Grevenmeyer, I., Kaul, N., Hauschild, J. & Pfender, M. Hydrothermal heat flux through aged oceanic crust: where does the heat escape? *Earth. Planet. Sci. Lett.* **202**, 159–170 (2002).

Acknowledgements This work was supported by the NSF, LANL/CULAR, and IGPP at UCSC.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.F. (e-mail: afisher@es.ucsc.edu).

Constant elevation of southern Tibet over the past 15 million years

Robert A. Spicer*, Nigel B. W. Harris*, Mike Widdowson*, Alexei B. Herman†, Shuangxing Guo‡, Paul J. Valdes§, Jack A. Wolfe|| & Simon P. Kelley*

* Department of Earth Sciences, The Open University, Walton Hall, Milton Keynes, MK7 6AA, UK

† Geological Institute, Russian Academy of Science, 7 Pyzhevskii Pereulok, 119017 Moscow, Russia

‡ Nanjing Institute of Geology & Palaeontology, Academia Sinica, Nanjing 210008, China

§ Department of Meteorology, Reading University, Earley Gate, PO Box 243, Reading RG6 6BB, UK

|| Desert Laboratory, Department of Geosciences, University of Arizona, Tucson, Arizona 85721, USA

The uplift of the Tibetan plateau, an area that is 2,000 km wide, to an altitude of about 5,000 m has been shown to modify global climate^{1–3} and to influence monsoon intensity^{4–8}. Mechanical and thermal models for homogeneous thickening of the lithosphere make specific predictions about uplift rates of the Tibetan plateau^{9,10}, but the precise history of the uplift of the plateau has yet to be confirmed by observations. Here we present well-preserved fossil leaf assemblages from the Namling basin, southern Tibet, dated to ~15 Myr ago, which allow us to reconstruct the temperatures within the basin at that time. Using a numerical general circulation model to estimate moist static energy at the location of the fossil leaves, we reconstruct the elevation of the Namling basin 15 Myr ago to be $4,689 \pm 895$ m or $4,638 \pm 847$ m, depending on the reference data used. This is comparable to the present-day altitude of 4,600 m. We conclude that the elevation of the southern Tibetan plateau probably has remained unchanged for the past 15 Myr.

Our present understanding of the uplift history of the Tibetan plateau is limited by the absence of a technique that can measure surface elevation directly because, unlike exhumation and erosion rates, which can be inferred from a range of well-calibrated techniques, palaeoaltitudes can be deduced only from proxy observations. In particular, there remains great uncertainty concerning the altitude of the Tibetan plateau during Neogene times, a critical period in the development of the monsoon^{11,12}, with consequent implications for Cenozoic global cooling. Elevation changes of the Tibetan plateau are driven ultimately by the Eocene collision¹³ and the continued convergence between the Indian and Eurasian plates. It has been argued from structural studies that the pre-collision surface of the southern plateau may have already achieved significant altitudes¹⁴.

A synthesis of Cenozoic deformation, magmatism and seismic structure suggests that the plateau of southern Tibet may have been uplifted during the Eocene¹⁵. The presence of high-K magmatism sourced from the subcontinental lithospheric mantle⁶ has been ascribed to delamination triggered by mantle convection^{3,9}. Extensional deformation, which is evident from north–south normal faults and regarded as indication of sufficient plateau uplift to result in east–west spreading³, has been dated at 7–9 Myr ago⁵, coeval with changes in foraminiferal populations in the Indian Ocean¹² and in the $\delta^{13}\text{C}$ signature and sediment production of Himalayan molasse^{4,11}, all of which have been cited as proxies for the intensity of the monsoon.

Other normal faults have yielded older ages (>13 Myr ago), however, which suggests that the southern and central Tibetan plateau has been an established geomorphological feature since at least the Middle Miocene^{16,17}, a conclusion supported by stable-

isotope palaeoaltimetry of Late Miocene (~10 Myr ago) north Himalayan carbonates¹⁸. It is also supported by a study of a north–south dyke swarm in southern Tibet, which established that east–west extension followed the generation of shoshonitic magmatism 18.3 ± 2.7 Myr ago (ref. 19). It is argued that by this time the southern Tibetan plateau had reached sufficient elevation to have attained excess potential energy after the convective removal of part of the subcontinental lithospheric mantle. But these proxies for altitude are strongly dependent on tectonic models of plateau formation.

Here we exploit the empirical relationship between leaf physiognomy and properties of the atmosphere related to altitude by using the climate leaf analysis multivariate program (CLAMP)^{20,21}. Although empirical, this relationship seems to be robust throughout the Tertiary²¹, in part because leaf physiognomy has a basis in convergent evolution constrained by the laws of physics and in part because the relationship between leaf morphology and climate is not subject to diagenetic modification. By applying the CLAMP methodology to Neogene Tibetan floras, we obtain the first estimate of palaeoaltitude for the Tibet plateau that is dependent on neither tectonic models nor stable-isotope systematics.

The Namling basin of southern Tibet (Fig. 1) comprises Neogene lava flows and shales with abundant plant-rich horizons interbedded with tuffaceous material unconformably overlying an Upper Cretaceous/Eocene volcanic basement. The western side of the basin has a series of steeply dipping, ash-rich, lacustrine sediments, within which two fossil leaf sites have been sampled.

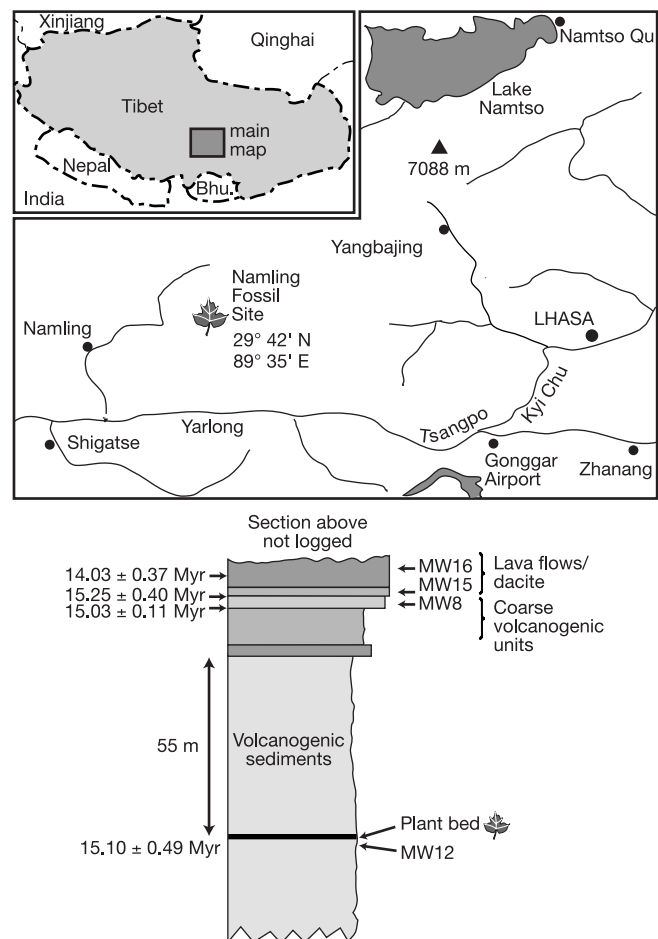


Figure 1 Upper part of the ~300-m logged section from Namling, indicating the stratigraphical position and ages of key dated samples. Inset shows location of Namling basin. MW indicates samples used in ^{40}Ar – ^{39}Ar analyses.

The lower site (locality A) lies at 4,300 m and the upper site at 4,600 m beneath capping lava flows at >4,800 m. At both localities abundant impressions in fine-grained water-lain grey tuffs represent leaves showing minimal mechanical fragmentation and no size sorting. We infer that the leaves were derived from trees growing proximal to the lake margins rather than from elevated sources on the distal edges of the basin.

Both localities yielded assemblages with similar floral composition. Taxa that normally rot quickly (*Alnus*, *Salix*) showed no sign of biodegradation, which suggests that the overall species composition is not biased because of preferential species decay. The age of flora from locality A (15.10 ± 0.49 Myr) was constrained by ^{40}Ar – ^{39}Ar dating of sanidine feldspar in a volcanogenic sediment lying immediately below it, and by phlogopite micas extracted from a potassic lava succession that overlies it (15.03 ± 0.11 Myr ago). Because both ages are within error, it is probable that the plant-bearing succession was deposited about 15 Myr ago.

The antiquity of the Namling flora precludes palaeoclimate, and the derived palaeoaltitude, being estimated from the modern climatic tolerances of the nearest living relatives (NLRs) of the ancient plants. This NLR technique is only useful for assemblages in which it is reasonable to expect modern species to occur. For 15-Myr floras, the constituent species are all extinct and can be compared only with their presumed modern descendants at genus level at best. Because the spread of environmental tolerances of species within a given genus is often large (for example, the genus *Salix* contains both tropical and Arctic species) the application of the NLR methodology in this instance involves large uncertainties. Alternative methods of better constrained precision need to be used to determine the palaeoaltitude of the Namling flora.

The only palaeobotanical technique that offers rigorous quantitative estimates of palaeoclimate and palaeoaltitude for leaf assemblages of this antiquity, and with quantifiable uncertainties, is one based on leaf physiognomy. The foliar physiognomy of woody dicots is demonstrably correlated with climate and, because this correlation is the product of the physical laws of gas diffusion, fluid flow and radiation balance, it is assumed to be stable over time^{20,21}. More than 400 collected specimens were categorized into 35 morphotypes using leaf architecture and venation characteristics and were subjected to a CLAMP analysis. CLAMP values for the Namling fossils yield a mean annual temperature (MAT) of $6.8 \pm 3.4^\circ\text{C}$ and a cold month mean temperature (CMMT) of $-6.2 \pm 5^\circ\text{C}$ using a data set containing the so-called ‘alpine nest’ samples (Physg3ar). An alternative data set (Physg3br) that lacks samples from cold environments gave an MAT of $8.1 \pm 2.3^\circ\text{C}$ with a

CMMT of $-4.19 \pm 3.7^\circ\text{C}$. The modest CMMTs suggest that either data set is appropriate for analysing the Namling fossils.

Although height can be calculated from lapse rates, in mountainous regions lapse rates are notoriously variable and dependent on local topography²². Because the ancient topography at the scale of individual depositional basins is unknown, we have used instead the more reliable approach of determining enthalpy, which is related to moist static energy (MSE). This thermodynamically conserved parameter of the atmosphere, which is related to both temperature and humidity, quantifies the total energy content of a parcel of air excluding the negligible kinetic energy^{21,23}. MSE is roughly conserved following an air parcel trajectory and, thus, in mid-latitudes it is roughly invariant with longitude²³. In lower latitudes, the longitudinal gradients can be larger. Because enthalpy is a function of temperature and moisture—two parameters that are crucial to plant growth—enthalpy can be decoded from foliar physiognomy using CLAMP^{21,23} (Fig. 2).

Analysis of the Namling assemblages using the Physg3ar data set yielded an enthalpy value of $289 \pm 7 \text{ kJ kg}^{-1}$, and the Physg3br set gave $290 \pm 6.4 \text{ kJ kg}^{-1}$. To determine altitude, the method requires an estimate of MSE at a known elevation. Normally this is based on one or more proximal and coeval sea-level floras from similar latitudes; however, such floras have not yet been recovered. To overcome this problem we calculated the Miocene MSE using the Hadley Centre numerical climate model. The advantage of this approach is that MSE can be calculated at the location of the fossil leaves and thus does not require any assumptions about the trajectories of the air parcels. Assuming longitudinal invariance was one of the largest sources of uncertainties²³ in previous North American studies, and such uncertainties increase for tropical latitudes.

The weakness of using the climate modelling approach lies in the uncertainties related to the modelling process. Although MSE is a relatively robust model variable, there are uncertainties arising from the representation of the physics in the model, and from our lack of knowledge of the detailed boundary conditions for the Miocene. Table 1 lists the potential sources of errors from this approach.

An estimate of the magnitude of errors associated with the model’s internal physics can be obtained from a comparison between the model-simulated modern climate and observations. Typical errors in tropical regions are of the order of 2 kJ kg^{-1} . This is a conservative estimate, however, because most climate models have been ‘tuned’ towards the observations. The effect of uncertainties in the Miocene boundary conditions can be estimated by performing sensitivity simulations. Changes to the height of Tibet do not significantly influence the model estimates (as expected from the theory²³). Errors related to palaeolongitude and palaeolatitude are relatively unimportant because we use the model predictions at the fossil sites (this would be the principal source of uncertainty if we were using sea-level fossil assemblages). The biggest uncertainty arises from lack of knowledge of the sea surface temperatures. We

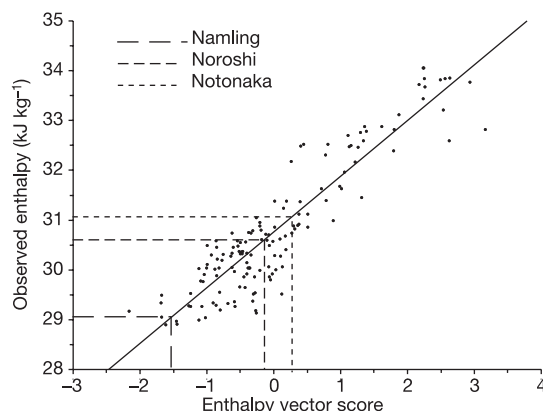


Figure 2 CLAMP enthalpy vector score against observed enthalpy regression for the Physg3br data set. Broken lines show the vector score and corresponding enthalpy predictions for the Namling flora and those from two coeval Miocene locations in Japan.

Table 1 Potential sources of errors in model simulations of moist static energy

Source of uncertainty	MSE uncertainty
Inaccuracies in simulating present-day distribution (calculated as the root mean square difference between the observations and the model in the region of Tibet)	2.3 kJ kg^{-1}
Palaeolongitude and palaeolatitude uncertainties	0.4 kJ kg^{-1}
Model sensitivity to variations in Tibetan elevation	Maximum of 0.6 kJ kg^{-1}
Changes in sea surface temperature	5.3 kJ kg^{-1} in the vicinity of Tibet; maximum of 8.2 kJ kg^{-1} over the globe

evaluated this uncertainty by performing sensitivity tests by prescribing a range of different sea surface temperatures, or ocean heat transports (when using a slab ocean model). Lack of a detailed palaeobathymetry precluded fully coupled ocean–atmosphere simulations.

To validate further the Miocene model simulations, we compared them with MAT and enthalpy values derived from the 15-Myr Notonakajima and Noroshi floras^{24,25} from Japan (mean MAT, 15 °C; mean enthalpy, 308 kJ kg⁻¹). The model sea surface temperatures were adjusted to ensure a good match to these sea level data (model-simulated MAT, 11 °C; mean enthalpy, 306 kJ kg⁻¹).

The model simulated Namling moist static energy is 335.5 kJ kg⁻¹ (see Supplementary Information). Changes in sea surface temperature typically gave variations of 5.3 kJ kg⁻¹ (in the Tibetan region, and we use this as our estimate of uncertainty) but elsewhere the variations could be as much as 8.2 kJ kg⁻¹. Overall, the height estimate for the Namling basin floor 15 Myr ago using the Physg3ar data set is given by:

$$Z = [(335.5 \pm 5.3) - (289.5 \pm 7.0)]/g = 46.0 \pm 8.78/9.81$$

$$= 4,689 \pm 895 \text{ m}$$

Using the Physg3br set, the palaeoelevation estimate is 4,638 ± 847 m. Including a worse-case modelling uncertainty of 8.2 kJ kg⁻¹ would give an uncertainty of 1,099 m. It should also be noted that there is a large seasonal cycle in enthalpy in the monsoon regions but all analyses have used the annual mean estimates.

These results have two important implications. First, the elevation of the valley floor has not changed significantly over the past 15 Myr. Any marked change in uplift rates that resulted from lithospheric delamination, as predicted by homogeneous thickening models, would have occurred before that time beneath southern Tibet. Second, if the plateau elevation had been stabilized at a threshold elevation above which it could not be supported by marginal stresses during the Early to Middle Miocene^{17,19}, then by 15 Myr ago the relief between plateau surface (~5,000 m) and intermontane basins was similar to the contemporary topography.

Our data provide a quantitative estimate of the altitude of southern Tibet 15 Myr ago; however, we caution against simple links being drawn between increases in the monsoon intensity (for example, at 8 Myr ago⁵ or 2.6 Myr ago⁷) and plateau uplift. Until a similar floral analysis is repeated for northern Tibet, no conclusions can be drawn about uplift of the whole plateau. Indeed, our results are consistent with geological arguments for a three-phase model of diachronous uplift with initial Eocene uplift in the south, uplift of central Tibet during the Oligo-Miocene and finally uplift of northern Tibet during the Plio-Quaternary¹⁵. The monsoon is driven, in part, by summertime sensible and latent heating effects of an uplifted plateau on the troposphere, and both plateau elevation and surface area determine the extent of these effects on atmospheric circulation^{1,3}. The complex interplay between elevation history and monsoon intensity can therefore be unravelled only once the increasing area of uplifted plateau throughout the Neogene is charted. Numerical climate-model experiments⁸ that assume uplift of the plateau evolved from small uplifted areas at moderate elevations (1,000–2,700 m) to large uplifted areas at high elevations (1,000–5,700 m) may therefore be unrealistic. The consequences of high elevations being achieved by progressively larger areas extending from south to north through time should now be examined. □

Methods

Dating

⁴⁰Ar–³⁹Ar analyses were done on mineral separates using the laser step-heating technique at the Open University.

CLAMP

Each morphotype was scored for 31 character states and evaluated by CLAMP using canonical correspondence analysis (CANOCO v.4.0) and Physg3ar and Physg3br

reference data sets for Northern Hemisphere (including Asia) temperate vegetation and climate (see Supplementary Information). Regression equations for CLAMP vector scores against observed climatic parameters were calculated in Axes 1–4 space. Uncertainties are here quoted as 2 s.d. of the residuals about the mean.

Enthalpy

We used the following moist static energy equation:

$$h = c_p T + L_v q + gZ = H + gZ$$

where h is moist static energy, c_p is the specific heat capacity of moist air at a constant pressure, T is temperature (in kelvin), L_v is the latent heat of vaporization, q is the specific humidity, g is gravitational acceleration and Z is the altitude. Enthalpy is given by $H = c_p T + L_v q$. The difference between two estimates of enthalpy for a given location yield an estimate of their difference in potential energy and hence height separation.

Received 17 October; accepted 9 December 2002; doi:10.1038/nature01356.

- Kutzbach, J. E., Guetter, P. J., Ruddiman, W. F. & Prell, W. L. The sensitivity of climate to late Cenozoic uplift in southern Asia and the American west: Numerical experiments. *J. Geophys. Res.* **94**, 18393–18407 (1989).
- Raymo, M. E. & Ruddiman, W. F. Tectonic forcing of late Cenozoic climate change. *Nature* **359**, 117–122 (1992).
- Molnar, P., England, P. & Martinod, J. Mantle dynamics, uplift of the Tibetan Plateau and the Indian monsoon. *Rev. Geophys.* **31**, 357–396 (1993).
- Burbank, D. W., Derry, L. A. & France-Lanord, C. Reduced Himalayan sediment production 8 Myr ago despite an intensified monsoon. *Nature* **364**, 48–50 (1993).
- Harrison, T. M., Copeland, P., Kidd, W. S. F. & Lovera, O. M. Activation of the Nyainqentanghla Shear Zone: implications for uplift of the southern Tibet Plateau. *Tectonics* **14**, 658–676 (1995).
- Turner, S. *et al.* Post-collision, shoshonitic volcanism on the Tibetan Plateau: implications for convective thinning of the lithosphere and the source of ocean island basalts. *J. Petrol.* **37**, 45–71 (1996).
- Qiang, X. K., Li, Z. X., Powell, C. McA. & Zheng, H. B. Magnetostratigraphic record of the Late Miocene onset of the east Asian monsoon and Pliocene uplift of northern Tibet. *Earth Planet. Sci. Lett.* **187**, 83–93 (2001).
- An, Z., Kutzbach, J. E., Prell, W. L. & Porter, S. C. Evolution of Asian monsoons and phased uplift of the Himalayan–Tibetan plateau since Late Miocene times. *Nature* **411**, 62–66 (2001).
- England, P. & Houseman, G. A. Extension during continental convergence, with application to the Tibetan Plateau. *J. Geophys. Res.* **94**, 17561–17579 (1989).
- Platt, J. P. & England, P. C. Convective removal of lithosphere beneath mountain belts: thermal and mechanical consequences. *Am. J. Sci.* **294**, 307–336 (1994).
- Quade, J., Rae, L., DeCelles, P. G. & Ojha, T. P. Development of Asian monsoon revealed by marked ecological shift during the latest Miocene in northern Pakistan. *Nature* **342**, 163–166 (1989).
- Kroon, D., Steens, T. & Troelstra, S. R. Onset of monsoonal related upwelling in the western Arabian Sea as revealed by planktonic foraminifers. *Proc. ODP. Sci. Res.* **117**, 257–263 (1991).
- Rowley, D. B. Age of initiation of collision between India and Asia: a review of stratigraphic data. *Earth Planet. Sci. Lett.* **145**, 1–13 (1996).
- Murphy, M. A. *et al.* Did the Indo-Asian collision alone create the Tibetan plateau? *Geology* **25**, 719–722 (1997).
- Tapponnier, P. *et al.* Oblique stepwise rise and growth of the Tibet Plateau. *Science* **294**, 1671–1677 (2001).
- Coleman, M. & Hodges, K. Evidence for Tibetan plateau uplift before 14 Myr ago from a new minimum estimate for east–west extension. *Nature* **374**, 49–52 (1995).
- Blisniuk, P. M. *et al.* Normal faulting in central Tibet since at least 13.5 Myr ago. *Nature* **412**, 628–632 (2001).
- Rowley, D. B., Pierrehumbert, B. S. & Currie, B. S. A new approach to stable isotope-based paleoaltimetry: implications for paleoaltimetry and paleohypsometry of High Himalaya since the Late Miocene. *Earth Planet. Sci. Lett.* **188**, 253–268 (2001).
- Williams, H. M., Turner, S., Kelley, S. & Harris, N. Age and composition of dikes in Southern Tibet: New constraints on the timing of east–west extension and its relationship to postcollisional volcanism. *Geology* **29**, 339–342 (2001).
- Wolfe, J. A. A method for obtaining climate parameters from leaf assemblages. *US Geol. Surv. Bull.* **2040**, 71 (1993).
- Wolfe, J. A., Forest, C. E. & Molnar, P. Paleobotanical evidence of Eocene and Oligocene paleoaltitudes in midlatitude western North America. *Geol. Soc. Am. Bull.* **110**, 664–678 (1998).
- Meyer, H. W. Lapse rates and other variables applied to estimating paleoaltitudes from fossil floras. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **99**, 71–99 (1992).
- Forest, C. E., Molnar, P. & Emanuel, K. A. Paleoaltimetry for energy conservation principles. *Nature* **374**, 347–350 (1995).
- Matsuo, H. in *Tertiary Floras of Japan* Vol. 1, *Miocene Floras* 219–243 (Geological Survey of Japan, Tokyo, 1963).
- Ishida, S. The Noroshi flora of Noto Peninsula, central Japan. *Mem. Faculty Sci. Kyoto Univ. Ser. Geol. Mineral.* **37**, 1–112 (1970).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank T. Tanai and K. Uemura for advice on the Japanese floras, and the National Environmental Research Council and Ferguson Fund for financial support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.A.S. (e-mail: R.A.Spicer@open.ac.uk).

Release of invasive plants from fungal and viral pathogens

Charles E. Mitchell & Alison G. Power

Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853, USA

Invasive plant species both threaten native biodiversity and are economically costly^{1–5}, but only a few naturalized species become pests^{2,4}. Here we report broad, quantitative support for two long-standing hypotheses that explain why only some naturalized species have large impacts. The enemy release hypothesis argues that invaders' impacts result from reduced natural enemy attack^{2,4,6–10}. The biotic resistance hypothesis argues that interactions with native species, including natural enemies, limit invaders' impacts^{6–8}. We tested these hypotheses for viruses and for rust, smut and powdery mildew fungi that infect 473 plant species naturalized to the United States from Europe. On average, 84% fewer fungi and 24% fewer virus species infect each plant species in its naturalized range than in its native range. In addition, invasive plant species that are more completely released from pathogens are more widely reported as harmful invaders of both agricultural and natural ecosystems. Together, these results strongly support the enemy release hypothesis. Among noxious agricultural weeds, species accumulating more pathogens in their naturalized range are less widely noxious, supporting the biotic resistance hypothesis. Our results indicate that invasive plants' impacts may be a function of both release from and accumulation of natural enemies, including pathogens.

Although the enemy release hypothesis (ERH) is commonly assumed to be a principal explanation for plant invasiveness^{4,7}, reviews have concluded that evidence for it is equivocal^{8,9}. Few data quantify the degree to which naturalized plants are released from natural enemies relative to their native range^{8–11}. In addition, a central prediction of the ERH—that naturalized plants that are more completely released from enemies have greater impacts—remains untested. Similarly, although some assumptions of the biotic resistance hypothesis (BRH) are supported^{8,11,12}, no study has tested whether naturalized plants experiencing greater attack by native enemies have lesser impacts. To test whether naturalized plants' release from and/or accumulation of pathogens explains their degree of impact, we compiled geographic data on plant associations with viruses and biotrophic foliar/floral fungi from published compendia and governmental or academic databases (see Methods). We focused on pathogens because they are ubiquitous and can be potent regulators of host populations^{10,13–15}.

Overall, each plant species was infected by 77% fewer fungus and virus pathogen species in its naturalized range than in its native range (general linear model: $F_{1,472} = 369.84$, $P < 0.0001$). We tested whether this decrease could be explained by changes in plant geographic range size¹¹, estimated by geographic prevalence. Pathogen species richness increased with greater plant geographic prevalence in both ranges, and more so in plants' native than in their naturalized ranges (range $\times$ geographic prevalence interaction: $F_{1,415} = 54.81$, $P < 0.0001$). In addition, the intercept of this relationship was also greater in the native range (native, 0.57 ± 0.14 (mean $\pm$ s.e.); naturalized, 0.012 ± 0.022). For all possible values of geographic prevalence ($0 \leq$ geographic prevalence ≤ 1), therefore, pathogen species richness was greater in plants' native ranges than in their naturalized ranges after controlling for the relationship between richness and range size.

We predicted viruses to be less easily escaped than fungi because they are more commonly systemic, seed-transmitted and asymptomatic than are fungi, and generally have broader host ranges¹⁶. As expected, plants were infected by 84% fewer fungi and 24% fewer viruses in their naturalized ranges than in their native ranges (Figs 1 and 2). Plants' greater release from fungi than viruses caused the composition of their pathogen assemblages to shift. The percentage of associated pathogen species that were fungi decreased from 88% in plants' native ranges to 61% in their naturalized ranges (Fig. 2; paired t -test: $t_{119} = 6.2$, $P < 0.0001$). We also predicted that plants heavily used by humans would be less released from pathogens because they probably have been introduced more times and at higher abundances than other species. Release from fungi was more

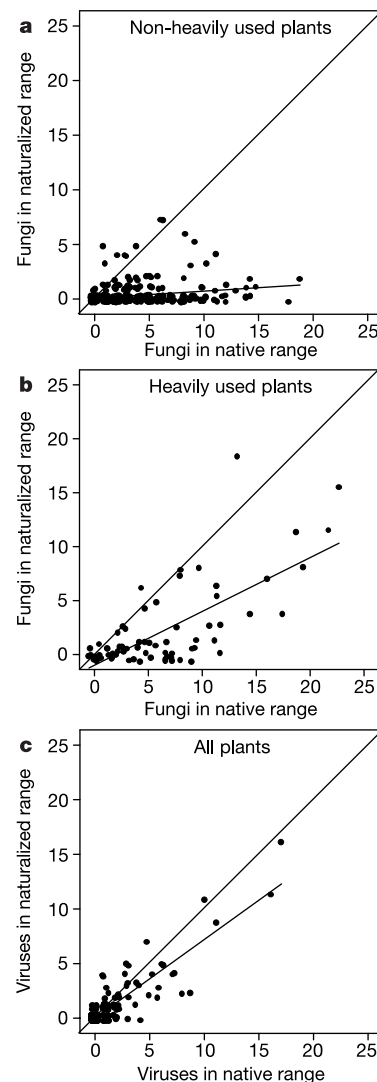


Figure 1 Net release of introduced plant species from fungal and viral pathogens. Each circle represents one host species; slight random noise was added to species with identical values to render each one visible. **a**, Release of the 401 plants not heavily used by humans from fungal pathogens. Number in naturalized range = $0.040 + 0.082 \times$ (number in native range); $\chi^2 = 62$, $P < 0.0001$, $R^2 = 0.135$. **b**, Release of the 72 plants heavily used by humans from fungal pathogens. Number in naturalized range = $0.042 + 0.34 \times$ (number in native range); $\chi^2 = 144$, $P < 0.0001$, $R^2 = 0.488$. **c**, Release of all 473 plants from viral pathogens. Number in naturalized range = $0.021 + 0.71 \times$ (number in native range); $\chi^2 = 164$, $P < 0.0001$, $R^2 = 0.809$.

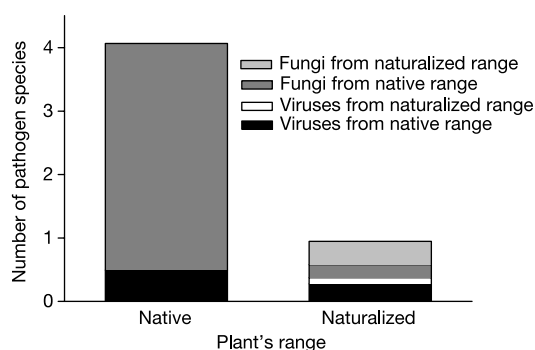


Figure 2 Composition of pathogen assemblages in the plant's native and naturalized ranges in terms of pathogen type and origin.

complete for plants not heavily used by humans than for heavily used species (Fig. 1a, b; Poisson ANCOVA: $\chi^2 = 73$, $P < 0.0001$), but release from viruses was not influenced by human use ($\chi^2 = 0.10$, $P = 0.75$). These patterns of pathogen release (Fig. 1) were general across plant species life histories, growth forms and phylogenetic histories. Controlling for host perenniality, growth form (graminoid, woody or herb), family (Asteraceae, Brassicaceae, Fabaceae, Poaceae or other) or monocot versus dicot did not increase the model fit by more than 2%.

Although the total costs and impacts of invasive species are relatively well quantified^{3,5}, attributing these totals to individual species remains challenging^{5,17}. We used plants' status as noxious weeds as an indicator of their economic costs because 95% of economic costs of US invasive plants are agricultural⁵ and states list plants as noxious weeds primarily on the basis of their agricultural impact¹. Many states and non-governmental organizations separately list invasive plants that degrade natural areas¹, providing an analogous indicator of plants' environmental impact. Among noxious weeds, species that experienced more complete pathogen release were more widely noxious (Fig. 3a), suggesting that greater release from pathogens increases invasive plants' economic costs,

and strongly supporting the ERH. Similarly, among plants listed as natural area invaders, species that experienced more complete pathogen release were more widely invasive (Fig. 3b). But this relationship was not robust to exclusion of the least released species (exclusion yielded, $\chi^2 = 3.0$, $P = 0.084$, $R^2 = 0.030$), providing more limited support for the ERH's prediction that more complete pathogen release increases invaders' environmental impact.

Pathogen release is the net result of two opposing forces: plants' 'escaping' pathogens from their native range during introduction, and subsequently 'accumulating' pathogens native to their naturalized range (release = escape - accumulation). Noxiousness might have increased with greater pathogen release (Fig. 3a) either because more noxious plants escaped more pathogens, or because they accumulated fewer pathogens, or both. Noxiousness increased with greater pathogen escape (Fig. 4a), implying that pathogens limit plant populations in their native range^{10,14,15} and supporting the idea that classical biocontrol can mitigate the costs of noxious weeds^{18–20}. In plants' naturalized ranges, fungi and viruses accumulated from those ranges comprised 49% of the pathogen assemblage (Fig. 2), indicating potential for biotic resistance despite plants' strong net release. Consequently, noxiousness also decreased with greater pathogen accumulation (Fig. 4b), supporting the BRH and the idea that native pathogens can be effective biocontrol agents^{12,18–20}. Thus, noxiousness is a function of both escape and accumulation of pathogens.

Control of potential pests by natural enemies is a chief ecosystem service²¹. Our results illustrate the breadth of this service: pathogens seem to both reduce the frequency with which naturalized plants are listed as noxious agricultural weeds (Fig. 4b) and prevent native plant populations from becoming pests (refs 14, 15, and Fig. 4a). Our results also suggest that invaders become more widely problematic, in part, because their introduction disrupts their associations with pathogens, thereby eroding the ecosystem service of pest control provided by pathogens and other natural enemies.

Together with similar results for animals²², our study shows that naturalized species are generally strongly released from pathogens. In addition, more completely released invasive plants are more

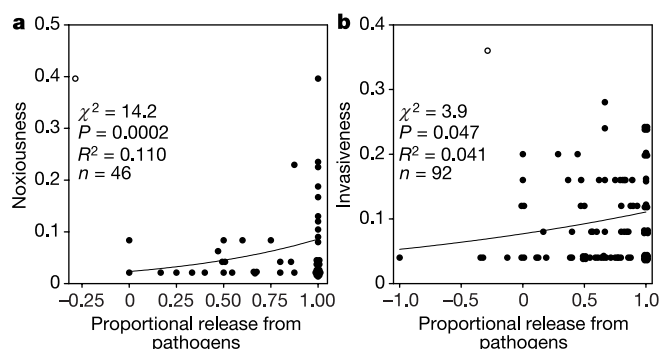


Figure 3 More complete release from pathogens increases the degree of noxiousness and invasiveness in the naturalized range. More complete release was measured by the difference in number of associated pathogen species between native and naturalized ranges divided by number associated in the native range. **a**, Degree of noxiousness in the naturalized range, measured by the proportion of states in which the pathogen is governmentally declared noxious; **b**, degree of invasiveness in the naturalized range, measured by the proportion of groups listing the species as an invasive problem in natural areas. Each circle represents one plant species; slight random noise was added to species with identical values to render each one visible. A highly influential outlier, *S. halepense* (Johnsongrass), was excluded from statistical analyses and is shown as an open circle (Methods).

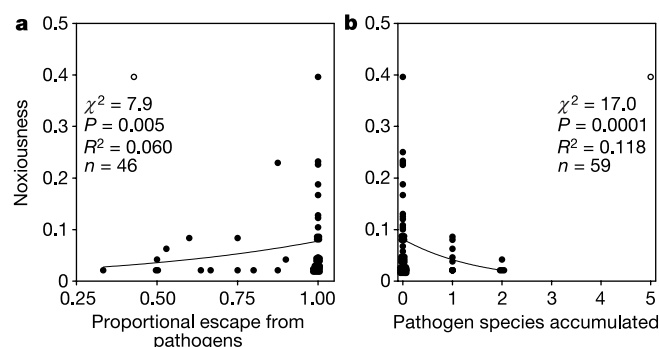


Figure 4 Decomposition of the effect of net release on noxiousness into the effects of escape from pathogens and accumulation of pathogens native to each plant's naturalized range. **a**, Escape from pathogens, measured as the proportion of pathogens from each plant's native range not occurring in its naturalized range; **b**, accumulation of pathogens native to each plant's naturalized range. The effect of pathogen accumulation was robust to removing all species that accumulated two pathogens from the analysis. Sample size is greater in **b** than in **a**, because plants with no pathogens reported in their native range, and therefore no potential for escape, were excluded from **a**. Each circle represents one plant species; slight random noise was added to species with identical values to render each one visible. A highly influential outlier, *S. halepense* (Johnsongrass), was excluded from statistical analyses and is shown as an open circle (Methods).

widely listed as problematic invaders of agricultural and natural ecosystems, strongly supporting the ERH. Although these correlations are not strongly predictive and exclude one outlier species, they were detected despite several limitations. We lacked data on pathogens' prevalences within populations, their distributions among plant populations within each geographic range, and the relative effects of different pathogens. We also lacked data on release from the natural enemies usually considered to be the principal agents of release and resistance for plants, herbivores^{8,9} and root pathogens^{10,14,15}, although release from these enemies could be correlated with release from pathogens quantified here. Therefore, additionally quantifying release from herbivores and other pathogens may explain much of the remaining variation in invasive plants' impacts. Our results point to a strong role for natural enemies in plant invasions; a more complete consideration of the diversity of invaders' natural enemies will better identify the extent of this role. □

Methods

Data collection

We focused on biotrophic foliar and floral fungi and all viruses of plants naturalized to the United States from Europe because this was the only combination of pathogens, plants and introduction route for which comprehensive data over two large geographic areas was available. In addition, plants naturalized to the United States from Europe are of interest in themselves because they are so numerous and important: they comprise 41% of the 4,100 plants naturalized to the United States and, on the basis of this percentage, have annual economic costs of US\$14 billion⁵. Similarly, the pathogens included comprise many of the world's most economically important groups¹⁶.

We compiled data on plant species in the US Department of Agriculture (USDA) Plants Database (<http://plants.usda.gov>) list of species naturalized to the United States (that is, those species not native to the United States but surviving in wild populations without human intervention). We used a random number generator to choose 1,165 of the 4,100 species listed on 17 January 2001. For each of these species, we determined whether its native range included part of Europe (species listed as from the 'Mediterranean' were included) using published floras from the naturalized range. For the 474 species native to Europe, we compiled data on growth form, perenniality and taxonomy from the USDA Plants Database. We excluded the one randomly chosen gymnosperm species, leaving 473 species in 69 families.

From the same database, we compiled the number of states in which each plant was governmentally declared a noxious weed and used this as an indicator of economic costs. For noxiousness, we report analyses including all states declaring a plant noxious, but results were similar when excluding prophylactic declarations of noxiousness. To estimate environmental damage resulting from each species, we counted the number of groups (state and regional exotic pest plant councils, state governmental agencies, the National Park Service and the Nature Conservancy) citing each species as an invasive problem in natural areas. These citations were compiled by the Alien Plant Working Group of the Plant Conservation Alliance, a consortium of US governmental agencies in cooperation with non-governmental groups (<http://www.nps.gov/plants/alien/list/all.htm>). Geographic prevalences in plants' native and naturalized ranges were defined as plants' proportional incidence in 39 European geographic regions²³ and in the 49 continental United States, respectively. Some species existing in naturalized populations are also cultivated by humans. We determined whether each species is or has been heavily used by humans (that is, a crop, silvicultural species or source of animal fodder) using ref. 24.

For each species, we counted the associated rust, smut and powdery mildew pathogens in the plant's native and naturalized ranges listed in the USDA Fungus-Host Distributions database (<http://nt.ars-grin.gov/SBMLweb/Databases/DatabaseHome.htm>). This database includes the data from numerous published compilations, including refs 25–27, and thus was regarded as a complete source of information on smut and powdery mildews in both Europe and the United States, and on rusts in the United States. The database did not include the authoritative compilations of European rusts^{28,29}, however, so we collected data from these texts directly. We counted the number of viruses of all types associated with each species in its native and naturalized range listed in the Virus Identification Data Exchange project database (<http://image.fs.uidaho.edu/vide>). We included only viruses known to infect each plant naturally, thereby excluding associations based only on experimental inoculations. We conservatively assumed that pathogens listed as occurring in both native and naturalized ranges followed the same path of introduction as their host, originating in the plant's native range.

Analysis

We used SAS Insight 8.0 for statistical analyses. Models with number of pathogens as the dependent variable were analysed assuming a Poisson response distribution and the identity link function, with two exceptions. The Poisson distribution could not be fit when testing whether number of pathogen species differed between native and naturalized ranges, and whether range size explained this difference, so a normal distribution was assumed. These two models treated species as block and range as the factor. Logit regressions were used for binomial response variables (noxiousness and invasiveness). All

χ^2 statistics are type III Wald tests. We report unadjusted R^2 values. All species were included in each analysis and figure unless otherwise stated. The range size analysis included all 420 species for which geographic prevalences were available. We considered that *Sorghum halepense* was an outlier (and thus excluded it) in analyses of noxiousness and invasiveness (Figs 3 and 4), because it has caused the most damage to US agriculture of plants in our dataset³⁰, and therefore the pathogens in its naturalized range have probably been more completely enumerated.

The observed release from pathogens could have resulted from pathogens being less studied in plants' naturalized than in their native ranges. This hypothesis predicts that better studied plant species, such as those governmentally declared noxious¹, should have more associated pathogen species. But plants designated as noxious weeds in at least one state had no more pathogens in their naturalized ranges than did other plants ($\chi^2 = 0.0035$; $P = 0.95$), suggesting that pathogen release was not an artefact of sampling effort.

Received 19 August; accepted 22 November 2002; doi:10.1038/nature01317.

1. US Congress Office of Technology Assessment *Harmful Non-Indigenous Species in the United States* Report No. OTA-F-565 (US Government Printing Office, Washington DC, 1993).
2. Williamson, M. *Biological Invasions* (Chapman & Hall, London, 1996).
3. Wilcove, D. S., Rothstein, D., Dubow, J., Phillips, A. & Losos, E. Quantifying threats to imperiled species in the United States. *BioScience* **48**, 607–615 (1998).
4. Mack, R. N. *et al.* Biotic invasions: causes, epidemiology, global consequences, and control. *Ecol. Appl.* **10**, 689–710 (2000).
5. Pimentel, D. (ed) *Biological Invasions: Economic and Environmental Costs of Alien Plant, Animal, and Microbe Species* (CRC, Boca Raton, Florida, 2002).
6. Darwin, C. *On the Origin of Species by Means of Natural Selection* (John Murray, London, 1859).
7. Elton, C. S. *The Ecology of Invasions by Animals and Plants* (Methuen, London, 1958).
8. Maron, J. L. & Vila, M. When do herbivores affect plant invasion? Evidence for the natural enemies and biotic resistance hypotheses. *Oikos* **95**, 361–373 (2001).
9. Keane, R. M. & Crawley, M. J. Exotic plant invasions and the enemy release hypothesis. *Trends Ecol. Evol.* **17**, 164–170 (2002).
10. Van der Putten, W. H. How to be invasive. *Nature* **417**, 32–33 (2002).
11. Clay, K. Correlates of pathogen species richness in the grass family. *Can. J. Bot.* **73**, S42–S49 (1995).
12. Hokkanen, H. Exploiter–victim relationships of major plant diseases: implications for biological weed control. *Agric. Ecosys. Environ.* **14**, 63–76 (1985).
13. Anderson, R. M. & May, R. M. Population biology of infectious diseases: Part I. *Nature* **280**, 361–367 (1979).
14. Packer, A. & Clay, K. Soil pathogens and spatial patterns of seedling mortality in a temperate tree. *Nature* **404**, 278–281 (2000).
15. Klironomos, J. N. Feedback with soil biota contributes to plant rarity and invasiveness in communities. *Nature* **417**, 67–70 (2002).
16. Agrios, G. N. *Plant Pathology* (Academic, San Diego, 1997).
17. Parker, I. M. *et al.* Impact: toward a framework for understanding the ecological effects of invaders. *Biol. Invasions* **1**, 3–19 (1999).
18. Julien, M. H. & Griffiths, M. W. (eds) *Biological Control of Weeds: A World Catalogue of Agents and their Target Weeds* (CABI, New York, 1998).
19. Charudattan, R. Biological control of weeds by means of plant pathogens: Significance for integrated weed management in modern agro-ecology. *BioControl* **46**, 229–260 (2001).
20. Liebman, M. in *Ecological Management of Agricultural Weeds* (eds Liebman, M., Mohler, C. L. & Staver, C. P.) 375–408 (Cambridge Univ. Press, Cambridge, 2001).
21. Naylor, R. L. & Ehrlich, P. R. in *Nature's Services: Societal Dependence on Natural Ecosystems* (ed. Daily, G. C.) 151–174 (Island, Washington DC, 1997).
22. Torchin, M. E., Lafferty, K. D., Dobson, A. P., McKenzie, V. J. & Kuris, A. M. Introduced species and their missing parasites. *Nature* (in the press).
23. Tutin, T. G. *et al.* (eds) *Flora Europaea* (Cambridge Univ. Press, Cambridge, 1964–1980).
24. Mabberly, D. J. *The Plant-Book: A Portable Dictionary of the Vascular Plants* (Cambridge Univ. Press, Cambridge, 1997).
25. Vánky, K. *European Smut Fungi* (G. Fischer, Stuttgart, 1994).
26. Farr, D. F., Bills, G. F., Chamuris, G. P. & Rossman, A. Y. *Fungi on Plants and Plant Products in the United States* (APS, St Paul, Minnesota, 1989).
27. Braun, U. *The Powdery Mildews (Erysiphales) of Europe* (G. Fischer, Jena, Germany, 1995).
28. Gaumann, E. A. *Die Rostpilze Mitteleuropas mit besonderer Berücksichtigung der Schweiz* (Buchdruckerei Böhler, Bern, 1959).
29. Wilson, M. & Henderson, D. M. *British Rust Fungi* (Cambridge Univ. Press, Cambridge, 1966).
30. Holm, L. G., Plucknett, D. L., Pancho, J. V. & Herberger, J. P. *The World's Worst Weeds: Distribution and Biology* (Univ. Hawaii, Honolulu, 1977).

Acknowledgements We thank B. Blossey, A. Dobson, A. Kuris, K. Lafferty, J. Maron, A. Rossman and M. Torchin for discussions; J. Gardell and B. Youn for assistance; and L. Aldrich-Wolfe, R. Allen, K. Blaisdell, G. Hall, K. Howe, G. Knight and S. Travers for manuscript reviews. This work resulted from the Disease and Conservation working group at the National Center for Ecological Analysis and Synthesis, which is supported by the US National Science Foundation (NSF), the State of California and the University of California Santa Barbara. Cornell University, the USDA and a postdoctoral research fellowship in microbial biology from the NSF also supported this work.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.E.M. (e-mail: cem46@cornell.edu).

Introduced species and their missing parasites

Mark E. Torchin*, Kevin D. Lafferty†, Andrew P. Dobson‡, Valerie J. McKenzie* & Armand M. Kuris*

* Marine Science Institute and Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, California 93106, USA

† US Geological Survey, Western Ecological Research Center, c/o Marine Science Institute, University of California, Santa Barbara, California 93106, USA

‡ Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544-1003, USA

Damage caused by introduced species results from the high population densities and large body sizes that they attain in their new location^{1–4}. Escape from the effects of natural enemies is a frequent explanation given for the success of introduced species^{5,6}. Because some parasites can reduce host density^{7–13} and decrease body size¹⁴, an invader that leaves parasites behind and encounters few new parasites can experience a demographic release and become a pest^{4,15}. To test whether introduced species are less parasitized, we have compared the parasites of exotic species in their native and introduced ranges, using 26 host species of molluscs, crustaceans, fishes, birds, mammals, amphibians and reptiles. Here we report that the number of parasite species found in native populations is twice that found in exotic populations. In addition, introduced populations are less heavily parasitized (in terms of percentage infected) than are native populations. Reduced parasitization of introduced species has several causes, including reduced probability of the introduction of parasites with exotic species (or early extinction after host establishment), absence of other required hosts in the new location, and the host-specific limitations of native parasites adapting to new hosts.

On average, 16 parasite species were recorded from native populations of host species. Of these, an average of only three parasite species successfully accompanied an invader to its introduced range. In addition, an average of four new 'native' parasites colonized the introduced host. In sum, introduced populations had roughly half the number of parasite species of native populations. These differences in parasite species richness between introduced and native ranges were significant when species richness was

standardized across studies (Figs 1 and 2a), and this effect was independent of sampling effort (Methods).

Introduced populations were also less heavily parasitized in terms of both average prevalence of each possible parasite species (4% in introduced versus 15% in native) and sum of the prevalences (71% in introduced versus 133% in native) of total parasite species per host population (where the prevalence of a parasite is the percentage of hosts that it infects in a population; Figs 1 and 2b, c). Average prevalence on a per-parasite-species basis (that is, parasites with zero prevalence were excluded from the calculation) did not differ between native and introduced populations (Figs 1 and 2d). In other words, parasites that invaded with their hosts achieved as high a prevalence in introduced populations (mean prevalence 28%) as in their native populations (mean prevalence 23%; Wilcoxon signed-rank test, $P_{\text{two-tailed}} = 0.18$). The parasite species left behind tended to be those that were less prevalent (mean prevalence 20%) in native populations as compared with those that did transfer (mean prevalence 27%; Wilcoxon sign-rank test, $P_{\text{one-tailed}} = 0.001$). For example, only the most prevalent of the seven reported trematode species that infect the snail *Batillaria cumingii* in its native range, Japan^{16–18}, has invaded the west coast of North America (M.E.T., J. Byers and T. Huspeni, manuscript in preparation). Native parasites that colonized introduced host populations (mean prevalence 29%) attained prevalences that were not significantly different from those introduced with the exotic host (mean prevalence 20%; paired $t = 0.31$, $P_{\text{two-tailed}} = 0.76$). Taken together, these findings suggest that there is nothing inherently different about the susceptibility of introduced populations versus native populations. Instead, parasites may be lost or 'filtered out' as a result of the invasion process.

Introduced populations are often derived from relatively small subsets of native populations (and sometimes from uninfected life-history stages), and this reduces the probability of introducing parasites along with a host species. Another potential limitation for the establishment of introduced parasites is that many parasites have complex life cycles requiring more than one host. If suitable hosts for all parasite life-cycle stages are not present, then the parasite will not become established. In addition, host population bottlenecks after introduction may break transmission of those parasites present in the founder population. For example, descendants of 100 adult European starlings, *Sturnus vulgaris*, released in New York City (1890–1891) spread over all regions of the United States^{15,19,20}. Of the 44 parasite species that we report from European starlings, a random sample of 100 invading birds should have had

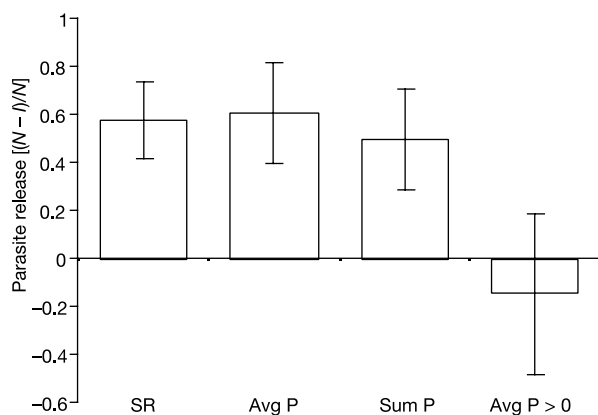


Figure 1 Parasite release experienced by introduced species. This release is represented by the proportion $(N - I)/N$, where N is the value for the native range and I is the value for the introduced range for standardized parasite species richness (SR), average prevalence (Avg P), summed prevalence (Sum P) and average prevalence on a per-

parasite-species basis (Avg P > 0; that is, parasites with zero prevalence were excluded from the calculation). This analysis is carried out on all taxa combined (based on data in Fig. 2). Error bars show 95% confidence intervals.

only 28 parasite species (estimated by an iterative resampling of 100 starlings from an infinite population, $n = 10,000$ trials, range = 19–37, with 95% of the time values falling between 23 and 33 parasite species). The small size of the founding starling population and lack of appropriate intermediate hosts might have further reduced the 28 expected parasite species to the nine species that we recorded in North American starlings. This is consistent with a previous quantitative study examining the parasites of native and introduced populations of starlings and house sparrows¹⁵.

Individuals arriving after an invader's population density increases could bring additional parasite species, and these would probably experience increased transmission efficiency at the higher host densities. For example, the black rat, *Rattus rattus*, was most certainly introduced repeatedly around the world. It is not surprising that, in our analysis, 38% (one of the highest values) of the rat's native parasites were also recovered from introduced populations.

Along with a complementary study of pathogens on introduced plant species²¹, to our knowledge this is the first taxonomically broad quantitative support, using a standardized analytical procedure, for the hypothesis that introduced species lose their native parasites and that their colonization by new parasites does not make up for that loss. Although the hypothesis that release from parasites may contribute to the success of an introduced species in a new environment is rarely examined quantitatively¹⁵, a study of the European shore crab, *Carcinus maenas*, shows that prevalences of parasitic castrators of the shore crab are negatively associated with demographic success (biomass and body size): introduced populations of *C. maenas* were not infected with these parasites and were significantly larger and had a greater biomass as compared with European populations¹⁴. Our results highlight the importance of

evaluating the role of parasites when examining the invasive species problem. More generally, invasions provide several opportunities to assess how parasites regulate host populations. In addition, their absence from introduced pest species suggests that the full potential of biological control to mitigate invasive species has not been explored as yet. □

Methods

Measures of parasitism

We analysed parasitological studies of 26 invasive species from seven taxa examined in their natural habitats (see Supplementary Information for full list of species and study selection criteria). To compare parasite measures across the diverse range of host taxa studied (which varied in their parasite richness), we standardized parasite species richness for each population of hosts in each study as a proportion relative to the total number of parasite species found in all studies in the native range of that host species. In addition, we compared both mean prevalence (averaged across parasite species for each host species and the summed prevalence (sum of all parasite species for each host species). The latter measure gives an indication of the unweighted cumulative extent of parasitism (or potential impact of parasitism on a host population) that each host experiences¹⁴. For each of these metrics, we estimated the proportional parasite release experienced by introduced species as $(N - I)/N$, where N is the value for the native range and I is the value for the introduced range of the above metrics.

Controls for potential confounds

We addressed two potential confounds of this approach. We expected to find a larger number of parasitological studies in native regions than in regions where the host had been introduced—an artefact that might lead to more comprehensive parasite lists in the native ranges and, therefore, could generate a spurious pattern with species richness consistent with our prediction (fortunately, prevalence is generally independent of sample size²²). However, a detailed analysis of the association between parasite species richness and number of hosts examined (in host's native ranges) showed that there was no significant association ($P > 0.05$) for all but four of the 26 species (*Bufo marinus*, $P = 0.02$; *Lepidodactylus lugubris*, $P = 0.02$; *Perca fluviatilis*, $P = 0.01$; and *Poecilia latipinna*, $P = 0.01$). In addition, in a general linear model with invasion status (either native or introduced) and host species as main effects, sample size was not significantly associated with parasite species richness ($P > 0.05$) and there was a significant effect of invasion status and host species on parasite species richness ($P = 0.0001$ and $P = 0.0001$, respectively). We also considered that a positive association between a species' geographical range and the community of parasites that it supports could confound our comparisons if species had limited introduced ranges relative to their native ranges. We controlled for this by averaging standardized parasite species richness (instead of summing standardized parasite species richness) across sample sites. This enabled us to use sites of relatively similar areas as replicates in native and introduced regions for each host species. In addition, introduced ranges were, on average, five times larger than native ranges, indicating that if such a bias existed, it ran counter to our results.

Received 20 August; accepted 29 November 2002; doi:10.1038/nature01346.

1. Vitousek, P. M. Biological invasions and ecosystem processes: Towards an integration of population biology and ecosystem studies. *Oikos* **57**, 7–13 (1990).
2. Wilcove, D. S., Rothstein, D., Dubow, J., Phillips, A. & Losos, E. Quantifying threats to imperiled species in the United States. *Bioscience* **48**, 607–615 (1998).
3. Ruiz, G. M., Fofonoff, P., Hines, A. H. & Grosholz, E. D. Non-indigenous species as stressors in estuarine and marine communities: Assessing invasion impacts and interactions. *Limnol. Oceanogr.* **44**, 950–972 (1999).
4. Torchin, M. E., Lafferty, K. D. & Kuris, A. M. Parasites and marine invasions. *Parasitology* **124**, S137–S151 (2002).
5. Keane, R. M. & Crawley, M. J. Exotic plant invasions and the enemy release hypothesis. *Trends Ecol. Evol.* **17**, 164–170 (2002).
6. Shea, K. & Chesson, P. Community ecology theory as a framework for biological invasions. *Trends Ecol. Evol.* **17**, 170–176 (2002).
7. Crofton, H. D. A model of host–parasite relationships. *Parasitology* **63**, 343–364 (1971).
8. Anderson, R. M. & May, R. M. Regulation and stability of host–parasite population interactions I. Regulatory processes. *J. Anim. Ecol.* **47**, 219–247 (1978).
9. May, R. M. & Anderson, R. M. Regulation and stability of host–parasite population interactions II. Destabilizing processes. *J. Anim. Ecol.* **47**, 249–267 (1978).
10. Scott, M. E. Regulation of mouse colony abundance by *Heligmosomoides polygyrus* (Nematoda). *Parasitology* **95**, 111–129 (1987).
11. Gulland, F. M. D. The role of nematode parasites in Soay sheep *Ovis aries* L. mortality during a population crash. *Parasitology* **105**, 493–503 (1992).
12. Kuris, A. M. & Lafferty, K. D. Modeling crustacean fisheries: Effects of parasites on management strategies. *Can. J. Fish. Aquat. Sci.* **49**, 327–336 (1992).
13. Hudson, P. J., Dobson, A. P. & Newborn, D. Prevention of population cycles by parasite removal. *Science* **282**, 2256–2258 (1998).
14. Torchin, M. E., Lafferty, K. D. & Kuris, A. M. Release from parasites as natural enemies: Increased performance of a globally introduced marine crab. *Biol. Invas.* **3**, 333–345 (2001).
15. Dobson, A. P. & May, R. M. in *Ecology of Biological Invasions of North America and Hawaii* (eds Mooney, H. A. & Drake, J. A.) 58–76 (Springer, New York, 1986).
16. Shimura, S. & Ito, J. Two new species of marine cercariae from the Japanese intertidal gastropod *Batillaria cumingii* (Crosse). *Jpn. J. Parasitol.* **29**, 369–375 (1980).
17. Rybakov, A. V. & Lukomskaya, O. G. On the life cycle of *Acanthoparyphium macracanthum* sp.n. (Trematoda, Echinostomatidae). **22**, 224–229 (1988).

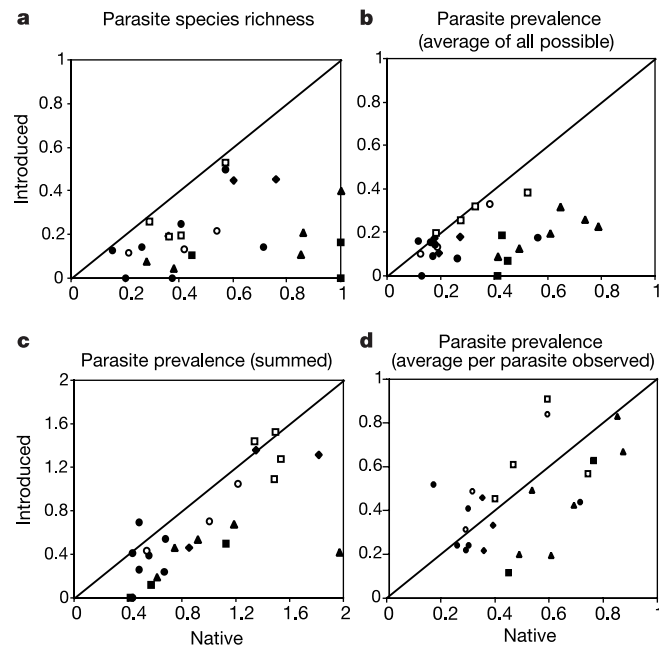


Figure 2 Parasitism in introduced and native populations. **a**, Standardized parasite species richness (estimated as $N_r = N_i/N_n$, where N_i is the number of parasite species found in the study and N_n is the total number of parasite species found in all studies in the native range of a given host species) in the native (x axis) and introduced (y axis) range. The diagonal line indicates no difference. **b**, Average prevalence (% of hosts infected, including parasites with zero prevalence). **c**, Summed prevalence (sum of prevalences of all parasite species). **d**, Average prevalence on a per-parasite-species basis (that is, parasites with zero prevalence were excluded from the calculation). In **a–d**, filled circles, molluscs ($n = 7$); filled squares, crustaceans ($n = 3$); filled triangles, fishes ($n = 6$); open circles, amphibians and reptiles ($n = 3$); filled diamonds, birds ($n = 3$); and open squares, mammals ($n = 4$).

18. Harada, M. & Suguri, S. Surveys on cercariae in brackish water snails in Kagawa Prefecture, Shikoku, Japan. *Jpn. J. Parasitol.* **38**, 388–391 (1989).
19. Chapman, F. M. *Handbook of Birds from Eastern North America* (Dover, New York, 1966).
20. Hair, J. D. & Forrester, D. J. The helminth parasites of the starling (*Sturnus vulgaris* L.): A checklist and analysis. *Am. Midl. Nat.* **83**, 555–564 (1970).
21. Mitchell, C. E. & Power, A. G. Release of invasive plants from fungal and viral pathogens. *Nature* **421**, 625–627 (2003).
22. Gregory, R. D. & Blackburn, T. M. Parasite prevalence and host sample size. *Parasitol. Today* **7**, 316–318 (1991).

Supplementary Information accompanies the paper on *Nature's* website
(<http://www.nature.com/nature>).

Acknowledgements This work was conducted as part of the Diseases and Conservation Biology Working Group supported by the National Center for Ecological Analysis and Synthesis, a centre funded by the National Science Foundation (NSF), the University of California, and the Santa Barbara campus. We thank S. Altizer, S. Gaines, P. Hudson, H. McCallum, A. W. Miller, C. Mitchell and A. Power for discussion and comments; A. Dove and G. Ruiz for providing data; and L. Mababa for data collection. This research was supported by NSF through the NIH/NSF Ecology of Infectious Disease Program, and by the National Sea Grant College Program, National Oceanic and Atmospheric Administration (NOAA), US Department of Commerce through the California Sea Grant College System, and in part by the California State Resources Agency. The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its subagencies. The US Government is authorized to reproduce and distribute for governmental purposes.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.E.T.
(e-mail: torchin@lifesci.ucsb.edu).

The contribution of *Shaker* K⁺ channels to the information capacity of *Drosophila* photoreceptors

Jeremy E. Niven^{*†}, Mikko Vähäsöyrinki^{†‡}, Mika Kauranen[‡],
Roger C. Hardie[§], Mikko Juusola^{*} & Matti Weckström[‡]

^{*} Physiological Laboratory, University of Cambridge, Cambridge CB2 3EG, UK
[‡] Department of Physical Sciences, Division of Biophysics, University of Oulu,
PO Box 3000, 90014 Oulun Yliopisto, Oulu, Finland
[§] Department of Anatomy, University of Cambridge, Cambridge CB2 3DY, UK
[†] These authors contributed equally to this work

An array of rapidly inactivating voltage-gated K⁺ channels is distributed throughout the nervous systems of vertebrates and invertebrates^{1–5}. Although these channels are thought to regulate the excitability of neurons by attenuating voltage signals, their specific functions are often poorly understood. We studied the role of the prototypical inactivating K⁺ conductance, *Shaker*^{6,7}, in *Drosophila* photoreceptors^{8,9} by recording intracellularly from wild-type and *Shaker* mutant photoreceptors. Here we show that loss of the *Shaker* K⁺ conductance produces a marked reduction in the signal-to-noise ratio of photoreceptors, generating a 50% decrease in the information capacity of these cells in fully light-adapted conditions. By combining experiments with modelling, we show that the inactivation of *Shaker* K⁺ channels amplifies voltage signals and enables photoreceptors to use their voltage range more effectively. Loss of the *Shaker* conductance attenuated the voltage signal and induced a compensatory decrease in impedance. Our results demonstrate the importance of the *Shaker* K⁺ conductance for neural coding precision and as a mechanism for selectively amplifying graded signals in neurons, and highlight the effect of compensatory mechanisms on neuronal information processing.

Insect photoreceptors have provided a model system for examining specific molecular mechanisms involved in information processing with graded voltage signals, including signal transduction (the phototransduction cascade)¹⁰ and membrane filtering (the photo-insensitive membrane)¹¹. Using these mechanisms, insect photoreceptors must compress the vast spatiotemporal range of light intensities to which they are exposed into voltage responses of limited amplitude and speed. In *Drosophila*, these mechanisms can be studied in relative isolation by patch-clamping dissociated photoreceptors, but *in vitro* photoreceptors do not survive prolonged light stimulation. By contrast, *in vivo* photoreceptors can be recorded intracellularly for more than an hour, and exposed to a full range of light intensities¹² (Fig. 1a). The photo-insensitive membrane of these cells contains three voltage-activated K⁺ channels: a *Shaker* channel that generates an A-type current, a slow delayed rectifier and, in some cells, a fast delayed rectifier⁹. The contribution of the *Shaker* K⁺ channel and its functional homologues (including vertebrate Kv channels)^{2,3} to neuronal function remains unclear, although they are thought to attenuate the amplitude of graded potentials and back-propagated action potentials in dendrites^{13–15}, to influence the firing frequency of spiking neurons¹⁶ and to determine the reliability of spike propagation¹⁷. The performance of a photoreceptor in coding a light signal can be described quantitatively by its sensitivity, signal-to-noise ratio and frequency response, allowing specific components of the signalling machinery, including ion channels, to be related to specific aspects of cellular

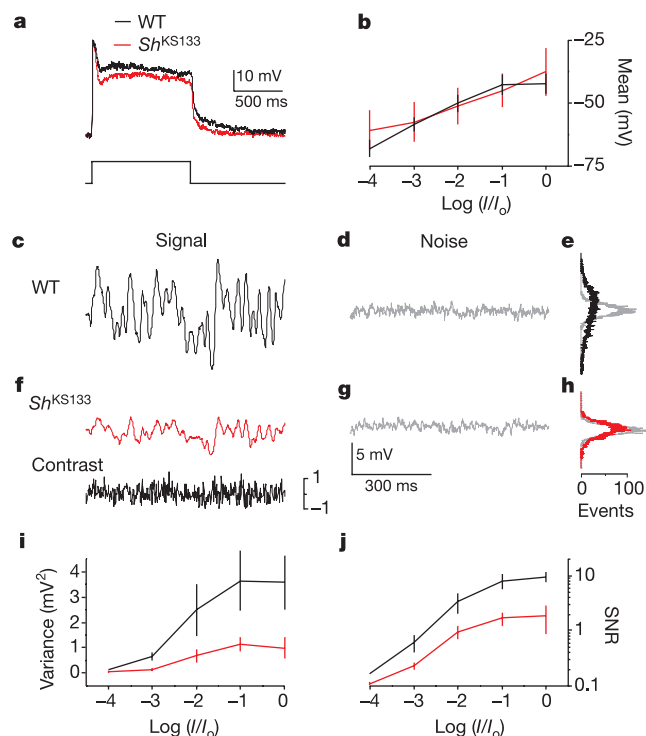


Figure 1 *Shaker* K⁺ channels amplify photoreceptor voltage responses. **a**, Responses of wild-type (WT, black) and *ShK133* (red) photoreceptors to a 1 s pulse of light. **b**, Mean (±s.e.m.) depolarization of WT (black) and *ShK133* (red) photoreceptors to dynamically modulated light contrast at five light intensities ($n = 6$ for each photoreceptor type in all experiments presented here). I , given background light intensity; I_0 , maximum background light intensity. **c, f**, Waveform of the average voltage signal of WT (black) and *ShK133* (red) photoreceptors to noise-modulated light contrast at the highest light intensity. **d, g**, Corresponding voltage noise (grey) for the averages presented in **c** and **f**. **e, h**, Distributions of the signal (WT, black; *ShK133*, red) and noise (grey) for **c–g**. **i, j**, The signal variance (**i**) and the signal-to-noise ratio (SNR, **j**) for WT (black) and *ShK133* (red) photoreceptors at each adapting-light background.

function¹¹. The information capacity of the photoreceptors, a measure of the number of states a signalling system can transmit in a given time window, can also be calculated from the signal-to-noise ratio¹⁸. We used these quantitative measures, combined with a mathematical model of the photoreceptors, to assess the contribution of the *Shaker* K^+ conductance to photoreceptor performance.

The signalling efficiency of photoreceptors in wild-type (WT) flies and *Sh^{KS133}* flies⁷, which produce non-functional *Shaker* K^+ channels, was studied by presenting sequences of dynamically modulated light with a mean contrast of 0.32, close to that of natural sceneries¹⁹, over the range of light intensities to which the photoreceptors are normally exposed (Fig. 1b). The light-induced current, assessed by the bump amplitude, quantum efficiency and macroscopic kinetics (see Methods), was unaffected in *Shaker* mutant photoreceptors. The photoreceptor signal response, $s(t)$, (Fig. 1c, f) and noise, $n(t)$, (Fig. 1d, g) were calculated at each light intensity²⁰ (see Methods). At all light background intensities, WT flies responded with a larger signal (measured as the signal variance) than *Sh^{KS133}* flies (Fig. 1i). There was no significant difference in the noise variance between *Sh^{KS133}* and WT flies over all light intensities ($P < 0.05$). Consequently, the signal-to-noise ratio, $SNR(t)$, of the WT flies was higher than that of the *Sh^{KS133}* flies over all light backgrounds except the dimmest (Fig. 1j), indicating that the *Shaker* K^+ channels amplify photoreceptor voltage responses.

How much better are WT photoreceptors in gathering information under dynamic, natural-like conditions than their *Sh^{KS133}* counterparts? To assess performance, we calculated the information capacity, C , of individual photoreceptors over a range of light intensities. Provided that the signal and noise are normally distributed (Fig. 1e, h), the information capacity of the photoreceptor in bits per second can be calculated from the frequency spectrum of the signal-to-noise ratio, $SNR(f)$ (refs 12, 18), using Shannon's formula²¹ (see Methods). The $SNR(f)$ increases with increasing light intensity, generating a concomitant increase in the photoreceptor

information capacity. A comparison of WT with *Sh^{KS133}* flies revealed that loss of the *Shaker* K^+ channels reduces the information capacity of the photoreceptors at all but the dimmest light levels, where noise dominates the responses of both cell types (Fig. 2a, b). This information loss was greatest over the lower frequency range of the spectrum (1–50 Hz), both photoreceptor types performing similarly at higher frequencies (50–150 Hz). The frequency distribution of the information depends solely on the $S(f)$ and $N(f)$. Because, under light-adapted conditions, the $N(f)$ of both WT and *Sh^{KS133}* photoreceptors behaved in a similar manner, the $S(f)$ determines the frequency distribution of the information. To examine the frequency-dependent amplification of the light-induced voltage signal, we calculated the photoreceptor frequency response function, $G(f)$, a measure of the contrast to voltage gain at each frequency²² (see Methods and Fig. 2c). Responses of WT and *Sh^{KS133}* photoreceptors show increased contrast gain and broadened bandwidth with increasing mean light intensity; however, the *Sh^{KS133}* responses contained more high-frequency signals (Fig. 2d).

To explain how loss of the *Shaker* conductance attenuates photoreceptor voltage responses, we developed a mathematical model of the photoreceptor that was based on Hodgkin–Huxley-type equations^{23,24}, which included *Shaker*, delayed rectifier and leak conductances (see Methods and Supplementary Information). The dynamic properties of the photoreceptor membrane were determined experimentally by injecting white-noise-modulated current. By driving the model with the same stimulus, it was possible to reproduce the WT and *Sh^{KS133}* membrane behaviour under both dark and light conditions in the time domain (see Methods and Fig. 3a, b). As a further comparison, we calculated first-order Wiener kernels, an approximation of the impulse response of the system, for the experimental and model WT and *Sh^{KS133}* membranes (Fig. 3c, d). Both the time domain simulations and kernel analysis suggested that the model incorporated the experimentally determined dynamic properties of these membranes. The experimental impedance function (resistance in the frequency domain)

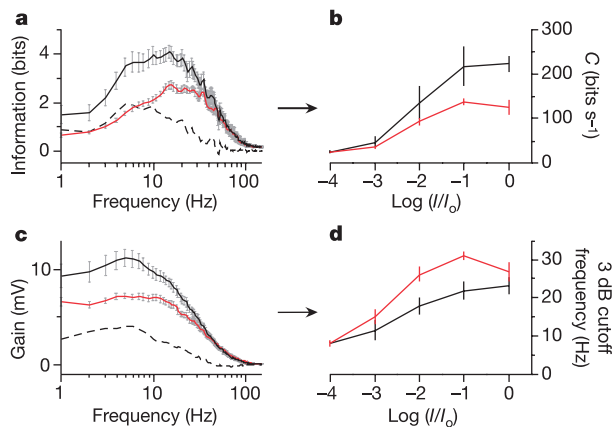


Figure 2 Comparison of the information and gain of wild-type and *Sh^{KS133}* photoreceptors. **a**, The mean ($\pm$ s.e.m.) frequency dependence of information extracted from the light-contrast stimulus in WT (black, $n = 21$) and *Sh^{KS133}* (red, $n = 13$) photoreceptors at background -1 . The dashed line is the difference between the WT and *Sh^{KS133}* photoreceptors. **b**, The information capacity, C (the integral of the information over all frequencies), at five adapting-light backgrounds for both *Sh^{KS133}* (red) and WT (black) photoreceptors. **c**, The mean ($\pm$ s.e.m.) frequency response function, $G(f)$, of the WT (black, $n = 21$) and *Sh^{KS133}* (red, $n = 13$) photoreceptors at background -1 . The dashed line is the difference in the gain of the WT and *Sh^{KS133}* photoreceptors. **d**, The 3 dB cutoff frequency (the point at which the gain falls to half the maximum) increases more rapidly in *Sh^{KS133}* (red) than in WT (black) photoreceptors as the light intensity increases.

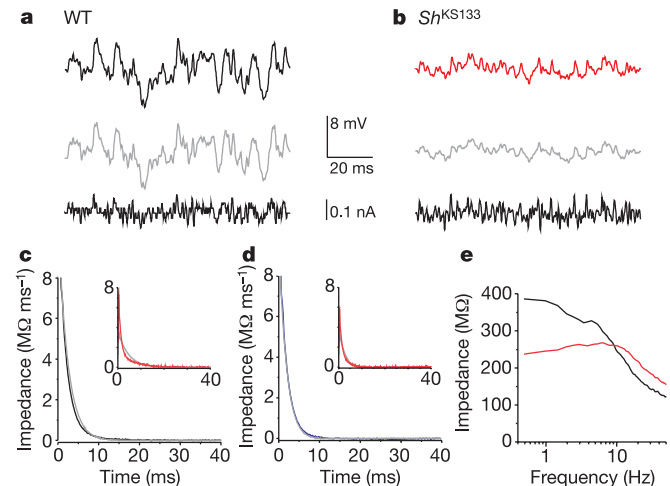


Figure 3 A photoreceptor model predicts accurately the dynamic responses of wild-type and *Shaker* membranes. **a**, Responses of light-adapted WT (upper) and model (middle) membranes to a dynamic current stimulus (peak-to-peak amplitude 0.1 nA) (lower). **b**, Responses of light-adapted *Shaker* (upper) and model (middle) membranes to a dynamic current stimulus (lower). **c**, **d**, First-order Wiener kernels for the WT and model membranes under dark (**c**) and light (**d**) adaptation. The inset shows first-order Wiener kernels for *Shaker* and model membranes under dark and light adaptation (axes are the same as those for WT kernels). **e**, Impedance functions of a WT (black) and a *Sh^{KS133}* (red) photoreceptor membrane. The *Sh^{KS133}* impedance function shows reduced gain at low frequencies relative to the WT impedance function.

showed lower impedance in the Sh^{KS133} membrane compared with the WT membrane, especially at low frequencies (Fig. 3e). These experimentally determined impedance functions showed similar characteristics to the frequency response function, $G(f)$ (Fig. 2c); the WT membrane had a higher gain than the Sh^{KS133} membrane at low frequencies.

Can the properties of the *Shaker* conductance account for the differences between frequency-dependent properties of WT and Sh^{KS133} photoreceptors? The overlap of the steady-state activation and inactivation functions of the *Shaker* K^+ channels means that a considerable fraction of these channels are open at the steady-state resting potential, creating a window current^{1,24} (Fig. 4a). In addition, the activation–inactivation functions of *Shaker* and delayed rectifier are separated on the voltage scale (Fig. 4a). These

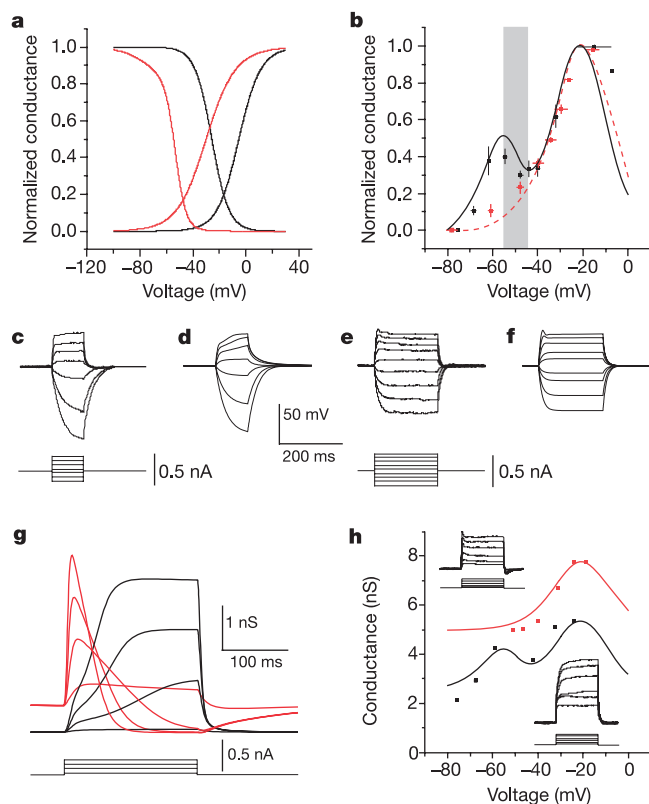


Figure 4 *Shaker*-mediated signal amplification is dependent on the activation–inactivation properties of the *Shaker* and delayed rectifier channels. **a**, Steady-state activation and inactivation curves for the *Shaker* (red) and delayed rectifier (black) conductances used in the model show the overlap of the activation and inactivation properties of the *Shaker* channels, creating a window current at rest. **b**, The normalized steady-state conductance ($\pm$ s.e.m.) of WT (black squares, $n = 6$) and *Shaker* (red squares, $n = 5$) photoreceptors as a function of membrane voltage, determined by injection of 200 ms current steps. Curves indicate the steady-state conductances predicted by the model for WT (black, solid) and *Shaker* (red, dash) photoreceptors. The shaded area shows the region during which signal amplification occurs due to the voltage separation of the *Shaker* and delayed rectifier channels. **c**, **e**, Intracellular voltage responses of *in vivo* WT (**c**) and Sh^{KS133} (**e**) photoreceptors to hyperpolarizing and depolarizing current pulses (maximum 0.38 nA). **d**, **f**, Model simulations of the WT (**d**) and Sh^{KS133} (**f**) voltage responses shown in **c** and **e**. **g**, Model simulations were used to resolve the behaviour of the *Shaker* (red) and delayed rectifier (black) K^+ conductances during current pulses in WT photoreceptors. During a current pulse, the *Shaker* conductance inactivates rapidly to below its initial level, boosting the voltage response. **h**, The total steady-state conductance of a WT (black squares) and a *Shaker* (red squares) photoreceptor. The curves show the model predictions of the total steady-state conductance for each photoreceptor. The inset shows the current steps from which the conductances were calculated: *Shaker* photoreceptor, upper left; WT photoreceptor, lower right.

two factors produce a voltage range, from -58 to -46 mV, in which the depolarization causes the total steady-state K^+ conductance to decrease (Fig. 4b) because *Shaker* channels inactivate but relatively few delayed rectifier channels are activated. Consequently, the *Shaker* conductance amplifies the voltage signal relative to a non-inactivating K^+ conductance. Is this amplification also present in photoreceptor responses to dynamic stimuli? WT photoreceptor voltage responses were gradually amplified during positive current steps, whereas those of Sh^{KS133} photoreceptors were attenuated (Fig. 4c, e). These voltage response characteristics were also successfully predicted by the model (Fig. 4d, f). To elucidate the amplification mechanism, we calculated the time courses of the voltage-activated K^+ conductances in the WT model responses to current steps (Fig. 4g). An increase in the photoreceptor voltage induced by current injection opens further *Shaker* channels that inactivate rapidly, thereby increasing the membrane resistance beyond the resting value, producing a larger voltage response to the same current step.

The frequency range, as seen in the impedance function (Fig. 3e), over which the amplification mechanism is effective is defined by the voltage-dependent inactivation time constant of the *Shaker* K^+ channel. In the voltage range of the photoreceptor responses to white noise current injection, this time constant is comparatively large (~ 40 – 100 ms), limiting the amplification to lower frequencies (see Supplementary Information). Therefore, loss of the *Shaker* conductance cannot fully explain the reduction in the Sh^{KS133}

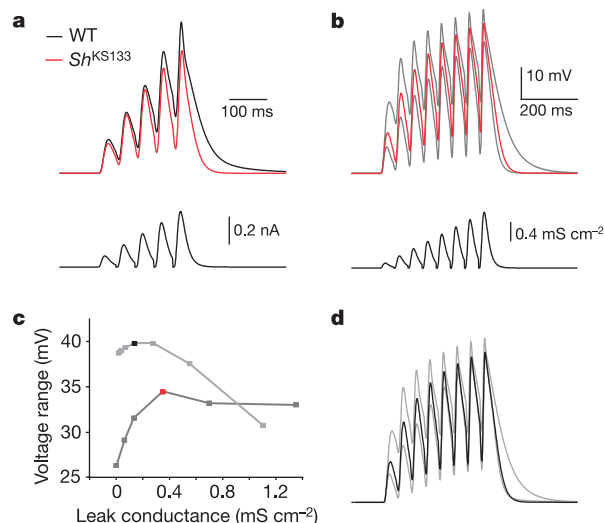


Figure 5 *Shaker* K^+ channel inactivation and the size of the leak conductance contribute to the voltage range of the WT and Sh^{KS133} photoreceptors. **a**, Normalized WT (black) and Sh^{KS133} (red) model responses to a sequence of five log-normal current pulses (below). **b**, Responses of the Sh^{KS133} model membrane (red) and two hypothetical model membranes (dark grey) to a series of log-normal conductance pulses (below). Each hypothetical membrane contained more or less leak conductance than the experimentally determined size of the leak conductance in the Sh^{KS133} photoreceptors. The range was taken as the difference between the minimum and maximum peak voltage responses to the conductance stimulus. Although the precise size of the range is specific for the conductance stimulus, the effects of varying the leak will be similar for all saturating stimuli. The stimulus was selected to resemble contrast changes typically observed in natural intensity time series. **c**, The size of the leak conductance in both the WT (grey) and *Shaker* (dark grey) models affects the size of the voltage range. The experimentally determined values for the WT (black square) and Sh^{KS133} (red square) photoreceptors maximize the available voltage range. **d**, Responses of the WT model membrane (black) and two hypothetical model membranes (grey) to a series of log-normal conductance pulses (below). Each hypothetical membrane contained more or less leak conductance than the experimentally determined size of the leak conductance in the WT photoreceptors.

photoreceptor impedance at higher frequencies. Indeed, the loss of the *Shaker* conductance would be expected to result in higher resistance at rest; however, Sh^{KS133} photoreceptors had reduced resistances compared with those of WT photoreceptors (Sh^{KS133} , $225.44 \pm 25.75 \text{ M}\Omega$ at $-64.3 \pm 5.4 \text{ mV}$, $n = 13$; WT, $410.53 \pm 34.45 \text{ M}\Omega$ at $-68.1 \pm 3.2 \text{ mV}$, $n = 21$). Current pulse experiments showed that the steady-state conductance ($1/R$, where R is resistance) of Sh^{KS133} photoreceptors behaved as would be expected of a WT membrane without *Shaker*, with no further voltage-dependent conductances (Fig. 4b, h). An additional leak is required to model the voltage-independent increase in the Sh^{KS133} conductance compared with that of the WT (Fig. 4h). This leak was larger ($\sim 2 \text{ nS}$) than the maximum steady-state *Shaker* conductance ($\sim 0.9 \text{ nS}$), but was of the same magnitude as the dynamic *Shaker* conductance in the mid-voltage range of the photoreceptor responses (Fig. 4g). Because of the size of the leak, it cannot be a pure K^+ conductance, which would hyperpolarize the membrane appreciably relative to the WT. To model the experimentally observed resting potential in Sh^{KS133} photoreceptors, two separate leak conductances were required that had reversal potentials of -85 mV (the K^+ equilibrium potential) and -30 mV (obtained by iteration), indicating a K^+ and a Cl^- conductance, respectively (data not shown).

How do the leak conductance and the *Shaker* conductance influence information processing in photoreceptors? The size of the leak conductance affects both the current-to-voltage gain (increasing leak reduces gain) and the spread of a signal across the voltage range. At rest, the *Shaker* conductance effectively acts as a leak conductance, reducing the current-to-voltage gain; however, at higher potentials, *Shaker* channel inactivation increases the gain. Thus, the *Shaker* conductance should increase the spread of the signal across the photoreceptor voltage range. To test this, we drove the model photoreceptors with a series of simulated pulses of light current¹² (Fig. 5a). The amplitudes of the voltage responses of the Sh^{KS133} photoreceptor were compressed compared with those of the WT, indicating that the WT photoreceptors were using the available voltage range more fully than Sh^{KS133} photoreceptors. This leads to a more efficient representation of the vast environmental light variations in the WT voltage range, improving their information processing. The amount of leak conductance cannot compensate for the loss of this amplification in Sh^{KS133} photoreceptors, but scales the amplitude of the voltage responses. Without either *Shaker* or additional leak conductances, photoreceptor voltage responses would also be saturating because of their high gain within a finite voltage range. By driving the model with conductance pulses of successively increasing amplitude, the effects of compression and saturation on the spread of the voltage signal could be seen (see Methods and Fig. 5b). Varying the size of the leak conductance in the model shows that the amount of this conductance present in Sh^{KS133} photoreceptors optimizes the available voltage range (Fig. 5c). Similarly, the amount of leak conductance in the WT photoreceptors is also the optimum for maximizing the voltage range (Fig. 5c, d). This suggests that leak conductance is adjusted in WT photoreceptors to maximize their voltage range and that the additional leak conductance in Sh^{KS133} photoreceptors enables them to compensate partially for the compression and saturation of their voltage range caused by loss of the *Shaker* conductance.

Our results indicate that a reduction in the information capacity of Sh^{KS133} photoreceptors can be explained by loss of the *Shaker* conductance and a compensatory increase in leak conductance. We show that the biophysical properties of *Shaker* channels lead to selective amplification, instead of attenuation^{1,14,15}, of the graded signals, and to efficient use of the available voltage range. These functions are likely to be widespread in processing graded signals in sensory receptors, and postsynaptic potentials in both vertebrates^{2–5,14–17} and invertebrates^{9,11,13,16}. In addition, the lower impedance of Sh^{KS133} photoreceptors suggests the presence of a tuning

mechanism that compensates the photoreceptors for loss of the *Shaker* conductance, highlighting the dynamic interaction between voltage-activated conductances, neuronal excitability and information coding. Although the details of such a compensation mechanism remain to be determined, there is increasing evidence that neurons may use homeostatic mechanisms to maintain their excitability^{25–27}. Such mechanisms are thought to compensate for changing synaptic inputs, but they may also be important for enabling sensory neurons to tune to environmental statistics during development. Our results show both the contribution of specific ion channel properties and the effects of compensatory mechanisms on neuronal information processing. □

Methods

Fly stocks

The WT strain was red-eyed *Drosophila melanogaster* Oregon Red. The null mutation in the *Shaker* channel, Sh^{KS133} (a missense mutation in the core region resulting in non-functional *Shaker* channels⁷), was also expressed in red-eyed flies. Both strains of fly were raised at 19°C in darkness.

Preparation and electrophysiology

In vivo intracellular recordings were carried out on photoreceptors from flies fixed in a custom-built holder¹². Recordings were made using quartz microelectrodes filled with 3 M KCl, with resistances between 150 and 220 M Ω . All recordings were made using a switch-clamp amplifier (SEC 10L, npi electronic) in current-clamp mode. The temperature of the flies was maintained at 25°C throughout the experiments. Photoreceptors were considered for analysis only if their membrane potential was less than -55 mV and they had at least a 45 mV saturating impulse response in dark-adapted conditions. Data acquisition, stimulus generation and signal analysis were carried out using a purpose-built MATLAB interface¹².

Whole-cell recordings of K^+ currents were made from photoreceptors of dissociated WT or Sh^{KS133} ommatidia^{8,9}. Bumps were elicited either by continuous dim illumination or by repeated brief flashes that contained, on average, less than one effective photon, and were detected and analysed off-line using MiniAnalysis software (Synaptosoft)²⁸.

Analysis

Single impaled photoreceptors were stimulated with repeated presentations of identical pseudorandom light contrast, $c(t)$ ($= \Delta I/I$, where ΔI is the change in intensity over time and I is the mean intensity over time), generated by a high-intensity green light-emitting diode subtending 1° (Marl Optosource) at five different intensity backgrounds over a range of more than 4 log units up to $3 \times 10^6 \text{ photons s}^{-1}$ (ref. 12). Averaging these responses gave the noise-free light-contrast photoreceptor voltage signal²⁹, $s(t)$, and subtraction of this signal from each individual trace gave the noise, $n(t)$. Dividing $s(t)$ by $n(t)$ gave the signal-to-noise ratio in the time domain, $\text{SNR}(t)$. The autospectra of $s(t)$ and $n(t)$ — $S(f)$ and $N(f)$, respectively—were calculated using a Fourier transformation. Dividing $S(f)$ by the corresponding $N(f)$ gave the signal-to-noise ratio in the frequency domain, $\text{SNR}(f)$. From the $\text{SNR}(f)$, the information capacity (bits s^{-1}), C , was calculated using Shannon's formula²¹:

$$C = \int_0^\infty (\log_2[\text{SNR}(f) + 1]) df$$

After averaging the stimulus and signal, the photoreceptor frequency response, $G(f)$, was calculated by dividing the cross-spectrum of the input (contrast) and output (photoreceptor signal) with the autospectrum of the input²². The frequency response, $G(f)$, was then expressed in terms of its gain, the ratio of the photoreceptor response amplitude to the stimulus amplitude¹².

To measure the photoreceptor impedance, a pseudorandom current waveform of small amplitude ($<0.2 \text{ nA}$) was injected through the electrode for 10 s to modulate the intracellular photoreceptor voltage. The photoreceptor impedance was calculated in the frequency domain between the average voltage response and the simultaneously recorded current stimulus waveform¹².

Model

A model of the photoreceptors was developed using MATLAB software (The MathWorks) on the basis of Hodgkin–Huxley-type equations²³. The model incorporated *Shaker* and slow delayed rectifier K^+ conductances (the fast delayed rectifier conductance was omitted because it is present in only some photoreceptors⁹), in addition to K^+ and Cl^- leak conductances. The voltage-dependent parameters (including time constants and steady-state functions for activation and inactivation) for these conductances were obtained either from published data^{8,9,29} or from further whole-cell patch-clamp experiments that were performed on isolated photoreceptors (see Supplementary Information). Other photoreceptor membrane properties (the maximum value of the active conductances, the resting potential, leak conductances and membrane capacitance) were estimated from *in vivo* recordings. The voltage-dependent properties of the ion channels, the reversal potentials for each ion and the membrane area were fixed parameters within the model, whereas other parameters were adjusted according to the experimental properties of each individual photoreceptor (see Supplementary Information).

Light channels were modelled as a leak conductance with an equilibrium potential of

+10 mV (see Supplementary Information). The size of the leak was adjusted to produce the same steady-state depolarization in the model as occurred in light-adapted photoreceptors (Fig. 1b). The log-normal shape of the light conductance pulses (Fig. 5) was fitted to experimentally derived light impulse responses¹², and the size of pulses was adjusted so that the largest conductance pulse produced a saturated voltage response with an amplitude similar to that seen in experiments.

To ensure that the Hodgkin–Huxley-type model could be driven with the same dynamic white noise that was used in experiments³⁰, we tested it by removing the active conductances, thereby reducing it to an analytically solvable RC circuit. A comparison of the model output with the exact solution of the RC circuit in the frequency domain showed that the numerical methods used in the model did not introduce errors. Identical stimuli were used in both modelling and *in vivo* recordings to allow their responses to be compared directly.

Received 1 November; accepted 2 December 2002; doi:10.1038/nature01384.

- Hille, B. *Ionic Channels of Excitable Membranes* 3rd edn (Sinauer Associates, Sunderland, Massachusetts, 2001).
- Rudy, B. Diversity and ubiquity of K⁺ channels. *Neuroscience* **25**, 729–749 (1988).
- Coetzee, W. A. *et al.* Molecular diversity of K⁺ channels. *Ann. NY Acad. Sci.* **868**, 233–285 (1999).
- Sheng, M., Liao, Y. J., Jan, Y. N. & Jan, L. Y. Presynaptic A-current based on heteromultimeric K⁺ channels detected *in vivo*. *Nature* **365**, 72–75 (1993).
- Wang, H., Kunkel, D. D., Martin, T. M., Schwartzkroin, P. A. & Tempel, B. L. Heteromultimeric K⁺ channels in terminal and juxtaparanodal regions of neurons. *Nature* **365**, 75–79 (1993).
- Salkoff, L. & Wyman, R. Genetic modification of potassium channels in *Drosophila* Shaker mutants. *Nature* **293**, 228–230 (1981).
- Kaplan, W. D. & Trout, W. E. The behaviour of four neurological mutants of *Drosophila*. *Genetics* **61**, 399–409 (1961).
- Hardie, R. C., Voss, D., Pongs, O. & Laughlin, S. B. Novel potassium channels encoded by the *Shaker* gene in *Drosophila* photoreceptors. *Neuron* **6**, 477–486 (1991).
- Hardie, R. C. Voltage-sensitive potassium channels in *Drosophila* photoreceptors. *J. Neurosci.* **11**, 3079–3095 (1991).
- Hardie, R. C. & Raghu, P. Visual transduction in *Drosophila*. *Nature* **413**, 186–193 (2001).
- Weckström, M. & Laughlin, S. B. Visual ecology and voltage-gated ion channels in insect photoreceptors. *Trends Neurosci.* **18**, 17–21 (1995).
- Juusola, M. & Hardie, R. C. Light adaptation in *Drosophila* photoreceptors: I. Response dynamics and signaling efficiency at 25°C. *J. Gen. Physiol.* **117**, 3–25 (2001).
- Laurent, G. Voltage-dependent nonlinearities in the membrane of locust nonspiking local interneurons, and their significance for synaptic integration. *J. Neurosci.* **10**, 2268–2280 (1990).
- Hoffman, D. A., Magee, J. C., Colbert, C. M. & Johnston, D. K⁺ channel regulation of signal propagation in dendrites of hippocampal pyramidal neurons. *Nature* **387**, 869–875 (1998).
- Magee, J., Hoffman, D., Colbert, C. & Johnston, D. Electrical and calcium signaling in dendrites of hippocampal pyramidal neurons. *Annu. Rev. Physiol.* **60**, 327–346 (1998).
- Connor, J. A. & Stevens, C. F. Voltage clamp studies of a transient outward membrane current in gastropod neural soma. *J. Physiol. (Lond.)* **213**, 21–30 (1971).
- Debanne, D., Guérineau, N. C., Gähwiler, B. H. & Thompson, S. M. Action-potential propagation gated by an axonal I_A-like K⁺ conductance in hippocampus. *Nature* **389**, 286–289 (1997).
- de Ruyter van Steveninck, R. R. & Laughlin, S. B. The rate of information transfer in graded-potential neurons and chemical synapses. *Nature* **379**, 642–645 (1996).
- Laughlin, S. B. A simple coding procedure enhances a neuron's information capacity. *Z. Naturforsch.* **36**, 910–912 (1981).
- Kouvvalainen, E., Weckström, M. & Juusola, M. Determining photoreceptor signal-to-noise ratio in the time and frequency domains with a pseudorandom stimulus. *Vis. Neurosci.* **95**, 1221–1225 (1994).
- Shannon, C. E. Communication in the presence of noise. *Proc. Inst. Radio Eng.* **37**, 10–21 (1948).
- Bendat, J. S. & Piersol, A. G. *Random Data: Analysis and Measurement Procedures* (Wiley & Sons, New York, 1971).
- Hodgkin, A. L. & Huxley, A. F. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol. (Lond.)* **117**, 500–544 (1952).
- Johnston, D. & Wu, S. M. *Foundations of Cellular Neurophysiology* (MIT Press, Cambridge, Massachusetts, 1995).
- Desai, N. S., Cudmore, R. H., Nelson, S. B. & Turrigiano, G. G. Critical periods for experience-dependent synaptic scaling in visual cortex. *Nature Neurosci.* **5**, 783–789 (2002).
- Stemmler, M. & Koch, C. How voltage-dependent conductances can adapt to maximize the information encoded by neuronal firing rate. *Nature Neurosci.* **2**, 521–527 (1999).
- Brickley, S. G., Revilla, V., Cull-Candy, S. G., Wisden, W. & Farrant, M. Adaptive regulation of neuronal excitability by a voltage-independent potassium conductance. *Nature* **409**, 88–92 (2001).
- Henderson, S. R., Reuss, H. & Hardie, R. C. Single photon responses in *Drosophila* photoreceptors and their regulation by Ca²⁺. *J. Physiol. (Lond.)* **524**, 179–194 (2000).
- Hevers, W. & Hardie, R. C. Serotonin modulates the voltage dependence of delayed rectifier and *Shaker* potassium channels in *Drosophila* photoreceptors. *Neuron* **14**, 845–856 (1995).
- Shampine, L. F. & Reichelt, M. W. The MATLAB ODE suite. *SIAM J. Sci. Comput.* **18**, 1–22 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank G. Garcia de Polavieja and H. Robinson for comments on an earlier version of this manuscript. The work was supported by the Royal Society (M.J.) and R.C.H.) and the Wellcome Trust (M.J., J.N. and R.C.H.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.J. (e-mail: mj216@cus.cam.ac.uk).

Ankyrin-B mutation causes type 4 long-QT cardiac arrhythmia and sudden cardiac death

Peter J. Mohler^{*†}, Jean-Jacques Schott[‡], Anthony O. Gramolini^{*}, Keith W. Dilly[§], Silvia Guatimosim[§], William H. duBell^{||}, Long-Sheng Song[§], Karine Haurogné[‡], Florence Kyndt[‡], Mervat E. Ali^{*}, Terry B. Rogers^{||}, W. J. Lederer[§], Denis Escande[‡], Herve Le Marec^{‡¶} & Vann Bennett^{*#}

^{*} Howard Hughes Medical Institute and Departments of Cell Biology, Biochemistry, and Neuroscience, Duke University Medical Center, Durham, North Carolina 27710, USA

[‡] Laboratoire de Physiopathologie et de Pharmacologie Cellulaires et Moléculaires, INSERM U533, Hôtel-Dieu; and [¶] Département de Cardiologie, Hôpital G&R Laennec, Nantes, France

[§] Medical Biotechnology Center and Department of Physiology, University of Maryland Biotechnology Institute; and ^{||} Department of Biochemistry & Molecular Biology, University of Maryland School of Medicine, Baltimore, Maryland 21021, USA

[†] These authors contributed equally to this work

Mutations in ion channels involved in the generation and termination of action potentials constitute a family of molecular defects that underlie fatal cardiac arrhythmias in inherited long-QT syndrome¹. We report here that a loss-of-function (E1425G) mutation in ankyrin-B (also known as ankyrin 2), a member of a family of versatile membrane adapters², causes dominantly inherited type 4 long-QT cardiac arrhythmia in humans. Mice heterozygous for a null mutation in ankyrin-B are haploinsufficient and display arrhythmia similar to humans. Mutation of ankyrin-B results in disruption in the cellular organization of the sodium pump, the sodium/calcium exchanger, and inositol-1,4,5-trisphosphate receptors (all ankyrin-B-binding proteins), which reduces the targeting of these proteins to the transverse tubules as well as reducing overall protein level. Ankyrin-B mutation also leads to altered Ca²⁺ signalling in adult cardiomyocytes that results in extrasystoles, and provides a rationale for the arrhythmia. Thus, we identify a new mechanism for cardiac arrhythmia due to abnormal coordination of multiple functionally related ion channels and transporters.

We previously characterized a large French kindred (Fig. 1a) where long-QT syndrome associated with sinus node dysfunction and episodes of atrial fibrillation segregated as an autosomal-dominant trait mapping to an 18-cM interval on chromosome 4q25–27 (ref. 3). Among the 25 affected patients (21 adults and 4 children) included in the study, average rate-corrected QT interval (QTc) was 490 ± 30 ms (for adults) and 465 ± 38 ms (for children) compared with 380 ± 30 ms and 403 ± 36 ms in unaffected individuals. T-wave morphologies characterized by sinusoidal features differed from those observed in the long-QT type 1–3 syndrome (LQT1–3). Sinus node bradycardia or junctional escape rhythm was diagnosed in all patients with LQT4 (ref. 3; see also Supplementary Fig. 1), although 24-h electrocardiogram (ECG) recordings revealed that sinus node dysfunction alternated with normal sinus rhythm. Nine patients were equipped with a rate-responsive atrial pacemaker because of marked bradycardia and the need of beta-blocking therapy. Finally, episodes of atrial fibrillation were diagnosed in 12 adult patients but were absent during childhood. Since the initial description of the family, eight additional individuals have been born. Four were demonstrated to carry the LQT4 haplotype. Sinus node abnormalities were diagnosed *in utero* in all affected members from generation IV.

Present address: Box 3892, Duke University Medical Center, Durham, North Carolina 27710, USA.

Sequencing of the gene encoding ankyrin-B (*ANKB*; also known as *ANK2*) identified an A to G transition mutation at position 4274 in exon 36, resulting in the substitution of glycine for a glutamic acid at amino acid residue 1425 (E1425G) near the regulatory domain of 220-kDa ankyrin-B (Fig. 1b). No nucleotide alterations were identified in two other positional candidate genes encoding CAMKII- δ or TRP3. Forty-five family members (24 carriers, including one individual who suffered sudden death, and 21 non-carriers) were evaluated for the E1425G mutation. The E1425G mutation segregated with LQT in 22 out of 24 individuals (III-5 and IV-1 were non-penetrant, with QTc = 420 ms), and with sinus node dysfunction in 23 out of 24 individuals (III-12 was non-penetrant with a heart rate of 60 beats per min (b.p.m.); Supplementary Fig. 4). The E1425G mutation was not found in more than 400 control alleles.

We evaluated the functional activity of the E1425G mutant on the basis of the ability to rescue abnormal Ca^{2+} dynamics of *AnkB*^{+/-} neonatal cardiomyocytes obtained from mice heterozygous for a null mutation in the gene encoding ankyrin-B^{4,5} (neonatal cardiomyocytes were used because adult cardiomyocytes are not readily transfected). Ankyrin-B expression in *AnkB*^{+/-} cells is reduced and

localized to a striated pattern only in certain regions of these cells (Supplementary Fig. 2). *AnkB*^{+/-} cardiomyocytes have a decreased spontaneous contraction rate (144 ± 10 to 78 ± 8 b.p.m.; $P < 0.05$) associated with prolonged intracellular Ca^{2+} concentration ($[\text{Ca}^{2+}]_i$) transients at a lower frequency (Fig. 1c; from 2.7 to about 1.3 Hz; $P < 0.05$). These defects in *AnkB*^{+/-} cardiomyocytes can be rescued by transfection with complementary DNA encoding green fluorescent protein (GFP)-tagged 220-kDa ankyrin-B (Fig. 1c; Ca^{2+} waves approximately 2.2 Hz, rhythm restored to 134 ± 11 b.p.m.). In contrast, *AnkB*^{+/-} cardiomyocytes transfected with ankyrin-B containing the human E1425G mutation still displayed abnormal Ca^{2+} oscillations (Fig. 1c; approximately 1.3 Hz; with instances of prolonged elevations in cytosolic Ca^{2+} ; $P < 0.05$) and a decreased beat frequency (71 ± 12 b.p.m.; $P < 0.05$), even though the mutant GFP-ankyrin-B itself targeted normally (Supplementary Fig. 2). Therefore, two normal copies of the ankyrin-B gene are required for normal Ca^{2+} signalling, and the E1425G mutation leads to loss-of-function. Ankyrin-B is, to our knowledge, the first identified protein to be implicated in a congenital long-QT syndrome that is not an ion channel or channel subunit^{1,6}.

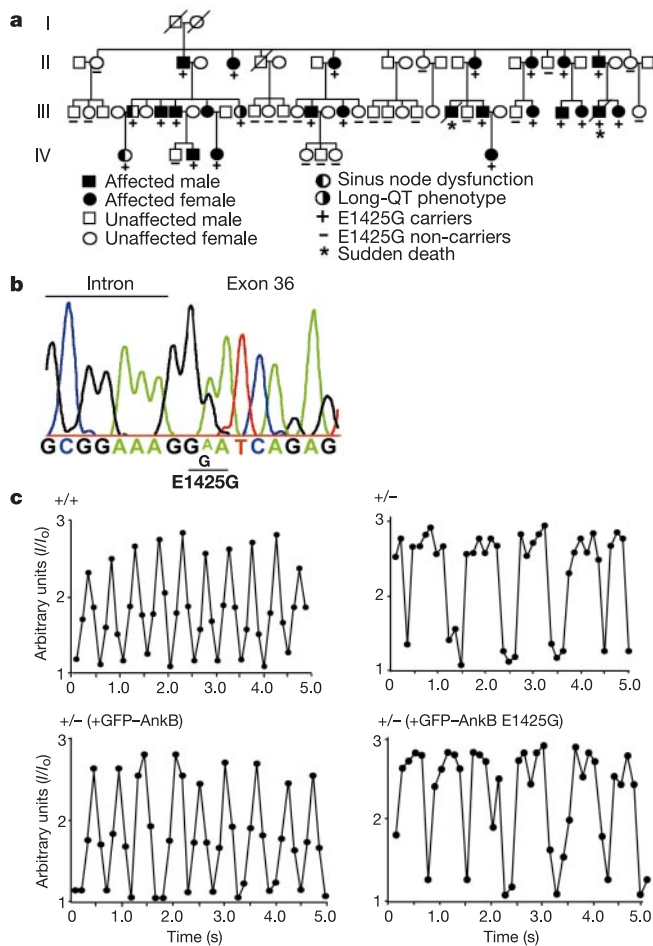


Figure 1 Loss-of-function mutation in ankyrin-B in type 4 long-QT syndrome. **a**, Pedigree of type 4 long-QT family. Filled symbols indicate long-QT and sinus node dysfunction phenotypes. **b**, An A to G mutation at position 4274 causes a E1425G missense mutation. **c**, Ca^{2+} levels as a function of time (fold increase over basal levels ($1/1$)). Graphs represent untransfected (+/+) (top left) and *AnkB*^{+/-} (+/-) neonatal cardiomyocytes (top right), and GFP-ankyrin-B (bottom left) and GFP-ankyrin-B E1425G (bottom right) transfected *AnkB*^{+/-} cardiomyocytes. After Ca^{2+} imaging, cardiomyocytes were monitored for GFP-ankyrin-B to ensure transfection.

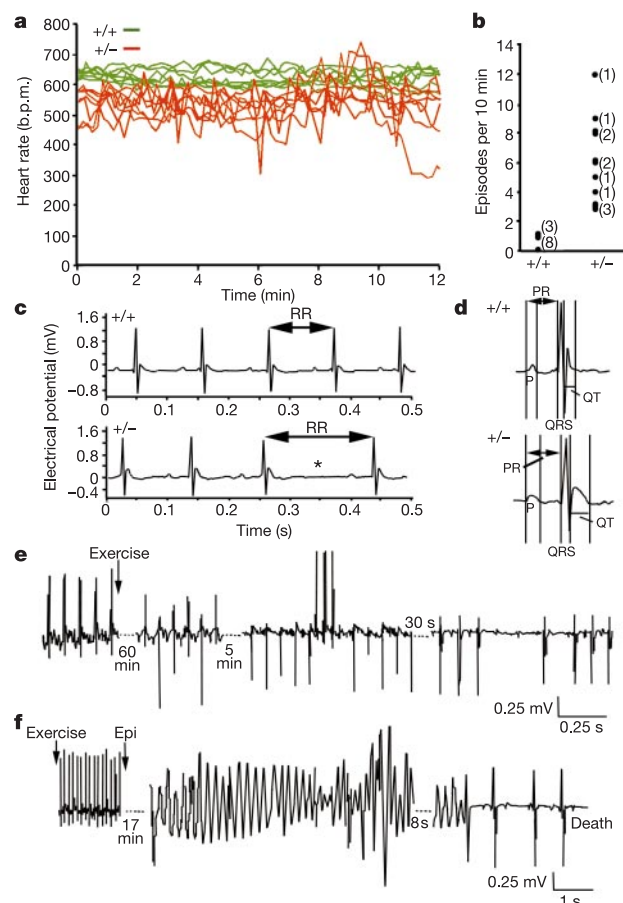


Figure 2 Sinus bradycardia, heart rate variability and sudden cardiac death in *AnkB*^{+/-} mice. **a**, Heart rates ($n = 7$ for wild-type and *AnkB*^{+/-} mice) showing bradycardia and variability in *AnkB*^{+/-} mice. **b**, Episodes of variable heart rate (greater than $\pm 10\%$ mean heart rate per animal) over 10 min ($n = 11$ for wild-type and *AnkB*^{+/-} mice). **c**, ECG of sinus slowing of a *AnkB*^{+/-} mouse. **d**, Sample ECGs of *AnkB*^{+/-} and wild-type mice. **e**, ECG traces of *AnkB*^{+/-} mice at rest and after exercise. **f**, ECGs after exercise and administration of epinephrine (Epi). Polymorphic ventricular arrhythmias occurred within about 17 min of epinephrine administration, followed by marked bradycardia and death 2 min after the arrhythmia. No wild-type mice exhibited changes in ECG patterns or died after these treatments.

Analysis of ECGs and heart rates of unrestrained animals using implanted radiotransmitter electrodes revealed significant similarities in cardiac phenotype between humans with LQT4 and *AnkB*^{+/-} mice (Fig. 2). *AnkB*^{+/-} mice have bradycardia with a conscious resting heart rate of 515 ± 49 b.p.m., compared with 641 ± 31 b.p.m. for wild-type mice (*n* = 12 for wild-type and 14 for *AnkB*^{+/-} mice; *P* < 0.05). Bradycardia was observed in all *AnkB*^{+/-} mice, with these mice displaying a heart rate of less than 600 b.p.m. for 87 ± 3.4% of a 30-min interval, whereas wild-type mice spend 4.3 ± 1.7% of the same interval at less than 600 b.p.m. (*n* = 10 for both genotypes; *P* < 0.05). *AnkB*^{+/-} mice also exhibit a high degree of heart-rate variability (Fig. 2a, b) associated with multiple episodes of abrupt sinus slowing. Figure 2c shows an ECG trace of one episode for a *AnkB*^{+/-} mouse. The prolonged RR intervals (sinus slowing; asterisk in Fig. 2c) occur on a background of reduced heart rate (Fig. 2a) compared with wild-type mice. Furthermore, *AnkB*^{+/-} mice exhibit episodes of intermittent isorhythmic atrio-ventricular dissociation similar to rhythm disturbances present in human LQT4 patients (Supplementary Fig. 3). ECG abnormalities in *AnkB*^{+/-} mice are not due to electrolyte or obvious structural defects in the heart, as no significant differences between wild-type and *AnkB*^{+/-} mice were evident in serum K⁺, Na⁺, Mg²⁺ or Ca²⁺, and no histopathological defects were detected in

heart sections from *AnkB*^{+/-} mice.

The QTc is significantly prolonged from 25 ± 1.0 to 30 ± 1.1 ms in *AnkB*^{+/-} mice (*n* = 9 and 11 for wild-type and *AnkB*^{+/-} mice; *P* < 0.05). The difference in apparent QT length in a mouse ECG could be due to delayed conduction and/or delayed repolarization⁷. ECGs of *AnkB*^{+/-} mice, in contrast to humans with LQT4, reveal general slowing of conduction with PR intervals increased from 35.9 ± 1.0 to 39.6 ± 0.7 ms, QRS intervals increased from 8.3 ± 0.1 to 11.2 ± 0.2 ms, and P-wave duration increased from 8.2 ± 0.7 to 13.4 ± 0.5 ms (*n* = 9 and 11 for wild-type and *AnkB*^{+/-} mice; all differences statistically significant, *P* < 0.05). Given that action potentials of adult *AnkB*^{+/-} cardiomyocytes are not substantially prolonged, the increase in QT interval observed in *AnkB*^{+/-} mice is probably due to delayed conduction.

Sudden cardiac death in humans with the E1425G mutation occurred after physical exertion and emotional stress (Fig. 1)³. We attempted to mimic these circumstances in mice with exercise followed by injection with epinephrine (see Methods). The mice responded in a dramatic manner. Two out of 14 *AnkB*^{+/-} mice became unresponsive for 3–10 s immediately after exercise alone. Over half of *AnkB*^{+/-} mice (8 out of 14) died after exercise combined with epinephrine. No wild-type mouse ever became unresponsive or died during these experiments (0 out of 6).

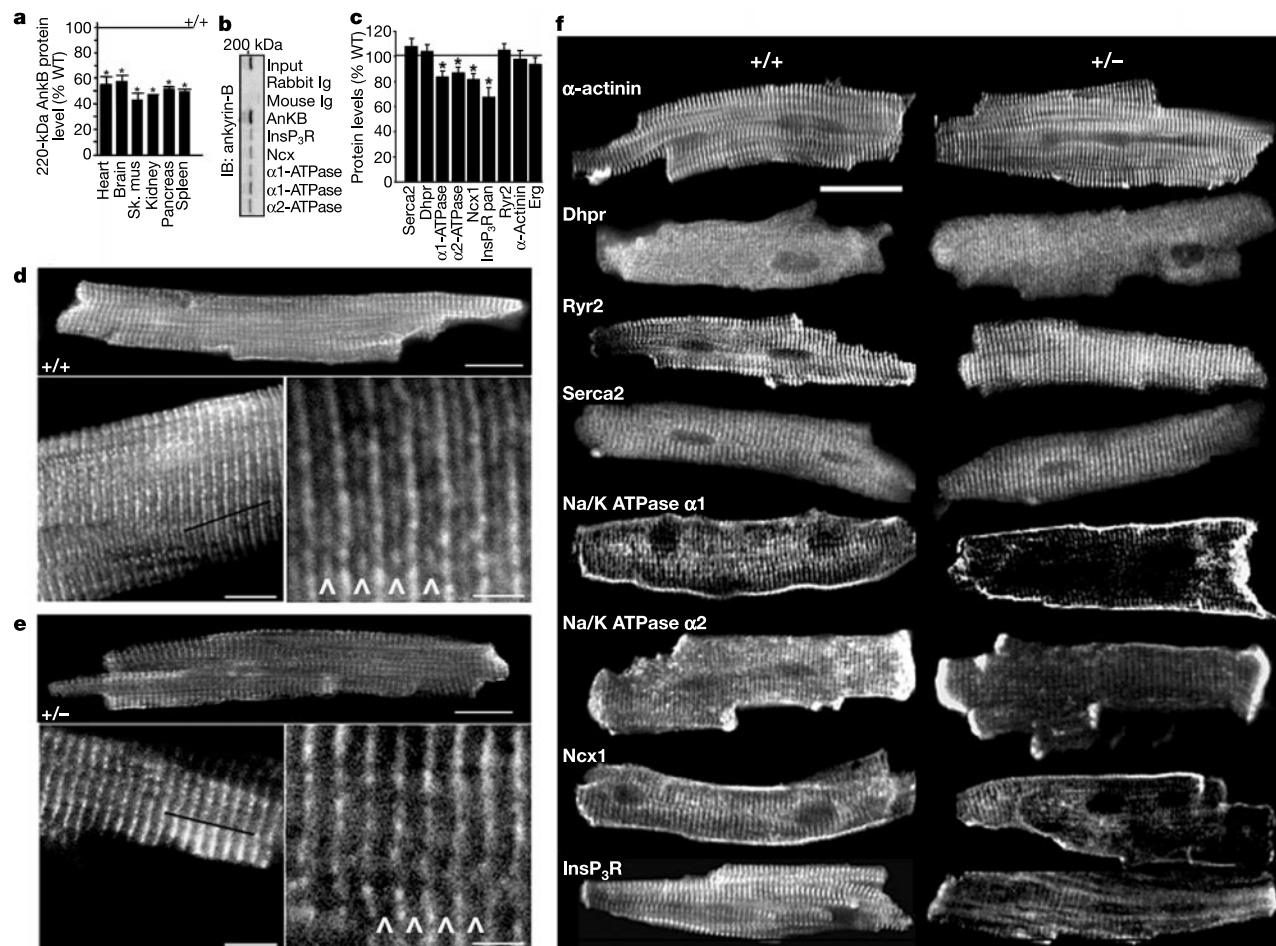


Figure 3 Coordinate reduction of ankyrin-B and ankyrin-B-associated proteins at Z-line/T-tubules of adult *AnkB*^{+/-} cardiomyocytes. **a**, Quantification of ankyrin-B protein expression in adult tissues (*n* = 4). **b**, Adult cardiomyocyte immunoprecipitations were analysed for 220-kDa ankyrin-B (input = 10% input). **c**, Quantitative immunoblots of protein levels in adult cardiomyocytes (*n* = 5). **d, e**, Ankyrin-B immunofluorescence in isolated wild-type and *AnkB*^{+/-} adult cardiomyocytes (arrowheads indicate

Z-lines/T-tubules; scale bars are 20 μm (top), 10 μm (bottom left) and 5 μm (bottom right). **f**, *AnkB*^{+/-} cardiomyocytes display qualitative loss of Ncx1, Na/K ATPase α₁ and α₂, and InsP₃R labelling over Z-line/T-tubules. Scale bar, 40 μm. Loss of ankyrin-B-associated protein staining at the Z-line/T-tubule was apparent throughout the entire depth of the cell.

Exercised *AnkB*^{+/-} mice displayed instances of reversed polarity of the QRS complex (two mice), and second-degree atrio-ventricular block (P-wave with no QRS complex, 11 mice; Fig. 2e). Prolonged polymorphic ventricular arrhythmia immediately preceding death was recorded in two mice treated with exercise plus epinephrine (Fig. 2f). The additional six mice that died from exercise and epinephrine displayed multiple short episodes (1–2 s) of polymorphic ventricular arrhythmia within 0–2 min before death. No arrhythmic episodes were observed in ECGs of wild-type mice after exercise or exercise plus epinephrine.

Reduction in the level of 220-kDa ankyrin-B by about 50% in immunoblots of adult cardiac tissue in *AnkB*^{+/-} mice (Fig. 3a) is accompanied by selective loss of ankyrin-B staining at the Z-line/transverse (T)-tubule region of *AnkB*^{+/-} cardiomyocytes (localization at the Z-line/T-tubule based on confocal Z-sections using dihydropyridine receptor (Dhpr) as a T-tubule marker; Fig. 3d). Ankyrin-B staining is retained at the M-line (predominant staining) and intercalated discs (Fig. 3e). Ankyrin-B is also aligned with Z-lines in skeletal muscle, but, in contrast to cardiomyocytes, ankyrin-B in skeletal muscle is restricted to costameres at the sarcolemma⁴.

The coordinate loss of ankyrin-binding proteins Na_v1.6, beta IV spectrin and neurofascin at axon initial segments lacking ankyrin-G⁸ suggested that reduced levels of ankyrin-B at Z-lines/T-tubules in heart could also result in the deficiency of proteins associated with

ankyrin-B at T-tubules. Na/K ATPase, Na/Ca exchanger (Ncx) and inositol-1,4,5-trisphosphate receptors (InsP₃R) are candidate ankyrin-binding proteins, on the basis of biochemical data⁹, that are localized at T-tubules¹⁰. Ankyrin-B co-immunoprecipitates with Ncx1, the α_1 and α_2 subunits of ATPase, and InsP₃R from extracts of heart tissue (Fig. 3b), but not with other cardiomyocyte proteins (including Dhpr, Serca2 and calsequestrin; data not shown). Immunoblots revealed that levels of InsP₃R (pan InsP₃R), α_1 and α_2 Na/K ATPase, and Ncx in isolated adult cardiomyocytes were reduced by 15–33% in *AnkB*^{+/-} cardiomyocytes (Fig. 3c). Measurements of the binding of [³H]InsP₃ (ligand for InsP₃R) and [³H]ouabain (ligand for Na/K ATPase) to adult cardiac microsomes also demonstrated a 33% and 16% reduction, respectively, in capacity in hearts of *AnkB*^{+/-} mice, whereas affinities for these ligands were unchanged. In contrast, quantitative western blot analysis revealed that protein levels of endoplasmic and sarcoplasmic reticulum components (Serca2, calreticulin, calsequestrin), K⁺ channels or associated subunits (Kcnq1/KvLqt1, Erg1, MinK/IsK, Kir2.1/Irk1, Kir2.3/Irk3), ryanodine receptor 2, plasma membrane Ca²⁺ channels (Dhpr, Pmca2), and structural proteins (α -actinin, dystrophin) were unaffected (Fig. 3c; see also Supplementary Fig. 6). Northern blots revealed no difference in the levels of mRNA encoding InsP₃R (type 1 and pan), Na/K ATPase α_1 and α_2 , and Ncx1.

The modest overall reduction in levels of Ncx, Na/K ATPase and InsP₃R has a substantial impact on the levels of these proteins localized at T-tubule sites, which can be detected easily by immunofluorescence (Fig. 3f). Ncx, as well as α_1 and α_2 Na/K ATPase are preferentially reduced in *AnkB*^{+/-} cardiomyocytes at T-tubule sites, whereas little change can be detected at the sarcolemma or intercalated discs. Furthermore, InsP₃R in *AnkB*^{+/-} cardiomyocytes is reduced at T-tubule sites and is disorganized in some regions, whereas label at intercalated discs is relatively normal. Markers for T-tubules (Dhpr), the sarcoplasmic reticulum (Serca2), and Z-line components (α -actinin)¹⁰ are unaltered in *AnkB*^{+/-} cardiomyocytes (Fig. 3f). Proteins that are also unaffected, as monitored by confocal analysis, include dystrophin, connexin 43, Na_v1.5, Na_v1.6, Erg1, Pmca2, Kcnq1/KvLqt1, calsequestrin and calreticulin (not shown). Ankyrin-B-dependent expression of InsP₃R, Ncx or Na/K ATPase may be a specialized feature of cardiac muscle, as there is no difference in expression or localization of these proteins in skeletal and vascular smooth muscle.

Reduction of Na/K ATPase, Ncx and InsP₃R in neonatal *AnkB*^{+/-} cardiomyocytes can be rescued by transfection with GFP-tagged 220-kDa ankyrin-B, but not by 220-kDa ankyrin-B containing the E1425G mutation (Supplementary Fig. 2). This indicates that 220-kDa ankyrin-B is necessary and sufficient for normal expression of Ncx, Na/K ATPase and InsP₃R in neonatal cardiomyocytes, and that the same E1425G mutation causing clinical arrhythmia in humans abolishes this activity. These findings establish that ankyrin-B participates in expression of multiple ion-channel/transporter proteins. The current list of ankyrin-B-dependent proteins may not be complete, as the full extent of ankyrin-B-binding partners is not yet known.

Examination of the electrical behaviour and Ca²⁺ dynamics of isolated heart cells from adult *AnkB*^{+/-} mice with ECG defects revealed a significant increase in the peak [Ca²⁺]_i level at all potentials (Fig. 4). No significant differences in resting levels of [Ca²⁺]_i (about 160 nM) were observed using indo-1. An increased [Ca²⁺]_i transient under these conditions implies that the amount of Ca²⁺ in the sarcoplasmic reticulum is elevated, although these values were not experimentally determined. No significant changes were evident in the magnitude or voltage dependence of the L-type Ca²⁺ channel current between -40 mV and +60 mV (*I*_{Ca}; Fig. 4a). Although heart weights were similar between wild-type and *AnkB*^{+/-} mice (indicating that no overt hypertrophy accompanied elevated [Ca²⁺]_i), cardiomyocytes of *AnkB*^{+/-} mice

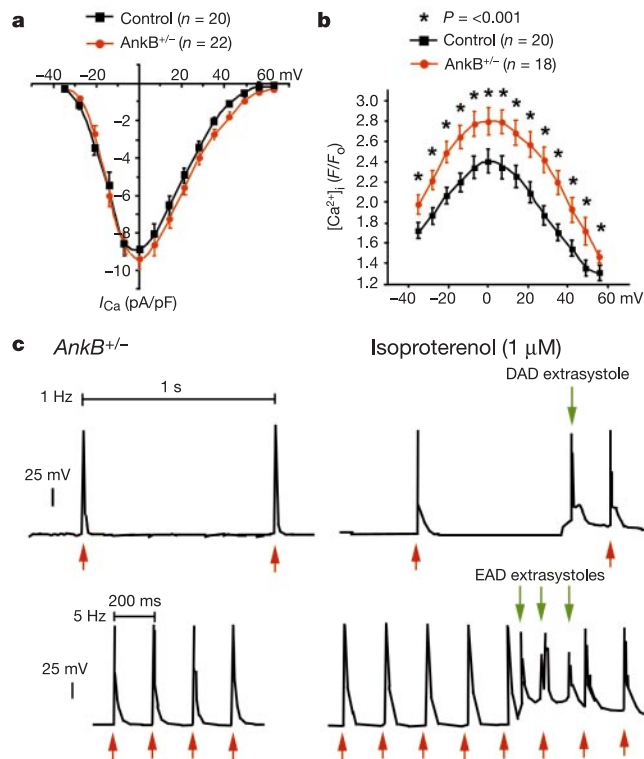


Figure 4 Ca²⁺ signalling in adult *AnkB*^{+/-} ventricular cardiomyocytes. **a**, *I*_{Ca} density (0 mV): *AnkB*^{+/-}, 9.37 ± 0.61 pA/pF; control, -8.87 ± 0.60 pA/pF, not significant. **b**, Peak of [Ca²⁺]_i transient plotted as *F*/*F*₀, where *F* is fluorescence intensity, and *F*₀ is resting fluorescence intensity. *F*/*F*₀ for *AnkB*^{+/-} is significantly greater than control at all membrane potentials. At 0 mV, *F*/*F*₀ for *AnkB*^{+/-} = 2.80 ± 0.13 (*n* = 18); control = 2.41 ± 0.12 (*n* = 20), representing a 16.2% increase. **c**, Steady-state action potentials recorded after 10–20 stimuli from *AnkB*^{+/-} cardiomyocytes at 1 and 5 Hz in control solutions and in isoproterenol. Red arrows indicate stimuli timing. DADs and EADs were observed in 36% of *AnkB*^{+/-} cardiomyocytes, but they were not observed in control cardiomyocytes.

did exhibit an approximately 23% increase in capacitance, suggesting increased surface area and a small increase in cell volume.

Cardiac action potentials measured in the presence and absence of isoproterenol revealed stress-induced abnormalities in *AnkB*^{+/-} heart cells (Fig. 4c). These cells were not significantly different from wild-type cells under control conditions (wild-type, AP₉₀ = 12.3 ± 1.0 (n = 8); *AnkB*^{+/-}, AP₉₀ = 15.0 ± 1.4 (n = 17), where AP₉₀ = time (in ms) for 90% repolarization of action potentials). However, after acute application of isoproterenol (1 μM) to simulate conditions of stress, action potentials in *AnkB*^{+/-} cardiomyocytes developed spontaneous extrasystoles at both 1 and 5 Hz, whereas control cells did not. Both delayed after-depolarizations (DADs) and early after-depolarizations (EADs) were observed in *AnkB*^{+/-} cells, and the DADs and EADs led to extrasystoles. The appearance of EADs, DADs and extrasystoles suggest that these arrhythmic mechanisms underlie the lethal arrhythmias seen in humans with the E1425G mutation, and might be caused by elevated [Ca²⁺]_i¹¹. A causative role of increased [Ca²⁺]_i in cardiac arrhythmia and congenital sudden cardiac death is an emerging area of interest, with current examples including gain-of-function mutations in the Ryr2 Ca²⁺-release channel¹².

Elevation in the [Ca²⁺]_i transient in *AnkB*^{+/-} cardiomyocytes can be rationalized by loss of Na/K ATPase isoforms^{13,14}. A small reduction of Na/K ATPase would be expected to mimic effects of cardiac glycosides such as digitalis, a Na/K ATPase inhibitor¹¹. Na/K ATPase inhibition results in increased [Ca²⁺]_i by first producing an increase in [Na⁺]_i, leading to a reduction of Ca²⁺ extrusion by Ncx^{13,14} into the extracellular space. In the face of unchanged Ca²⁺ entry by I_{Ca}, and combined with a small reduction in Ncx, the reduction of Na/K ATPase in *AnkB*^{+/-} cells should lead to an increase in total cellular Ca²⁺, as suggested by our data. Long-term reduction or increase in Ncx in animal models does not seem to produce a severe phenotype because of diverse compensatory mechanisms¹⁵. Therefore the loss of Na/K ATPase is probably the major contributor to elevated [Ca²⁺]_i transients in *AnkB*^{+/-} ventricular myocytes. This study shows that ankyrin-B has an important involvement in regulating the coordinated expression of Ncx, the Na/K ATPase, InsP₃R, and possibly other ankyrin-binding proteins. Moreover, the ankyrin-B dysfunction in humans leads to lethal cardiac arrhythmias and type 4 long-QT syndrome. □

Methods

Human mutation analysis

Genomic DNA was prepared from peripheral blood lymphocytes. Mutation analysis was conducted by direct sequencing of the ankyrin-B gene. All 45 exons of the ankyrin-B gene were amplified using intronic primers (Supplementary Fig. 5).

Mouse ECG recordings

We performed and analysed mouse ECG recordings as described in the Supplementary Information.

AnkB^{+/-} cardiomyocytes

GFP-ankyrin-B E1425G was created using standard molecular techniques. Neonatal cardiomyocytes were prepared, transfected³ and imaged using Fluo3-AM⁴. Transfection was optimized for low GFP-ankyrin-B expression (levels undetectable without GFP-labelled antisera⁵). After Ca²⁺ imaging, cells were immunostained using indicated antisera and visualized using Alexa 568 so that signal would not interfere with Fluo3 fluorescence. For contraction experiments, over 100 cells were monitored for each condition. For rescue experiments, Ca²⁺ was monitored in over ten cells for each condition.

Immunoblotting and immunoprecipitations

We performed quantitative immunoblots using equal protein concentrations as described⁴. Adult heart immunoprecipitations were performed using standard techniques (lysis buffer: 1.5% Triton X-100, 0.5% deoxycholate plus × 2 protease inhibitor cocktail, most proteins were soluble except for InsP₃R (40%)). Immunoblotting was performed using ¹²⁵I-labelled protein A, and intensities were quantified by phosphorimaging⁵.

Immunostaining and imaging

Wild-type and *AnkB*^{+/-} cells were prepared and imaged identically as described⁵. Antibodies used were: α-actinin, dystrophin and Dhpr (Sigma), InsP₃R types 1 and 2

(ABR), pan InsP₃R (Calbiochem), Pmca2, Ryr2 and Serca2 (ABR), GFP (Chemicon or Clontech), Ncx1 (RDI), Na/K ATPase α1 (Developmental Studies Hybridoma Bank; Upstate Biotechnology; M. Caplan) α2 (Upstate Biotechnology), Erg1 and connexin 43 (Chemicon), Nav1.6, Kir2.1, Kir2.3 and MinK (Alomone), Nav1.5 (W. Catterall), Kcnq1 (KvLqt1; Santa Cruz), and ankyrin-B monoclonal and affinity purified polyclonal immunoglobulin. Similar results were obtained in both isolated cardiomyocytes and in sections of adult cardiac muscle.

Patch-clamp methods

AnkB^{+/-} and wild-type animals (1–3 months of age) were killed by an intraperitoneal injection of pentobarbital sodium (100 mg per kg). Single cardiomyocytes were isolated¹⁶. An Axopatch-200A or -200B amplifier (Axon Instruments) was used to measure membrane currents^{16,17}. The patch pipette (1–3 MΩ) solution was (in mM): CsCl (130), NaCl (10), MgATP (5), HEPES (10), MgCl₂ (1), TEA-Cl (20) pH 7.2 (with CsOH). Superfusion solution 1 contained (in mM): NaCl (140), KCl (5), MgCl₂ (0.5), CaCl₂ (1.8), NaH₂PO₄ (0.33), glucose (5.5) and HEPES (5), pH 7.4, at 35–37 °C. Superfusion solution 2 was the same as solution 1 but with CsCl substituted for KCl + 10 μM TTX. After conversion to whole-cell voltage clamp in solution 1, solution 2 was used to measure I_{Ca}. Test depolarizations followed 4 50-ms depolarizations to 0 mV at 1 Hz. A 500-ms ramp-depolarization from -90 mV to -40 mV was followed by a 50-ms period at -40 mV before test depolarizations.

Action potential recordings

Cardiomyocytes were superfused with solution 1. β-adrenergic stimulation of cells was produced by the addition of 1 μM isoproterenol to solution 1. Pipette filling solution was as above, except that KCl was substituted for CsCl (pH 7.2 with KOH). Axopatch 200A was used in current clamp mode to record action potentials. Current injections triggered action potentials at a constant rate (1 Hz, 5 Hz). We performed all experiments at 37 °C.

Confocal [Ca²⁺]_i imaging

Biorad MRC600 and Zeiss LSM510 microscopes were used with simultaneous electrical measurements to determine [Ca²⁺]_i (refs 16, 18). In parallel experiments, measurements of resting [Ca²⁺]_i were obtained by adding indo-1 (25 μM)¹⁹ to the pipette filling solution, on a system made by the authors. Resting [Ca²⁺]_i was calculated as described¹⁹.

Statistics

Data were analysed using either paired two-tailed *t*-tests or two-way analysis of variance, and *P* values <0.05 were considered significant. Data are expressed as means ± s.e.m.

Received 31 July; accepted 5 November 2002; doi:10.1038/nature01335.

- Keating, M. T. & Sanguinetti, M. C. Molecular and cellular mechanisms of cardiac arrhythmias. *Cell* **104**, 569–580 (2001).
- Mohler, P. J., Gramolini, A. O. & Bennett, V. Ankyrins. *J. Cell Sci.* **115**, 1565–1566 (2002).
- Schott, J. J. *et al.* Mapping of a gene for long QT syndrome to chromosome 4q25–27. *Am. J. Hum. Genet.* **57**, 1114–1122 (1995).
- Tuvia, S., Buhusi, M., Davis, L., Reedy, M. & Bennett, V. Ankyrin-B is required for intracellular sorting of structurally diverse Ca²⁺ homeostasis proteins. *J. Cell Biol.* **147**, 995–1008 (1999).
- Mohler, P. J., Gramolini, A. O. & Bennett, V. The Ankyrin-B C-terminal domain determines activity of Ankyrin-B/G chimeras in rescue of abnormal inositol 1,4,5-trisphosphate and ryanodine receptor distribution in ankyrin-B (-/-) neonatal cardiomyocytes. *J. Biol. Chem.* **277**, 10599–10607 (2002).
- Towbin, J. A. & Vatta, M. Molecular biology and the prolonged QT syndromes. *Am. J. Med.* **110**, 385–398 (2001).
- Casimiro, M. C. *et al.* Targeted disruption of the Kcnq1 gene produces a mouse model of Jervell and Lange-Nielsen Syndrome. *Proc. Natl Acad. Sci. USA* **98**, 2526–2531 (2001).
- Jenkins, S. M. & Bennett, V. Ankyrin-G coordinates assembly of the spectrin-based membrane skeleton, voltage-gated sodium channels, and L1 CAMs at Purkinje neuron initial segments. *J. Cell Biol.* **155**, 739–746 (2001).
- Bennett, V. & Baines, A. J. Spectrin and ankyrin-based pathways: metazoan inventions for integrating cells into tissues. *Physiol. Rev.* **81**, 1353–1392 (2001).
- Frank, J. S. & Garfinkel, A. in *The Myocardium* (ed. Langer, G. A.) 1–32 (Academic, San Diego, California, 1997).
- Bers, D. M. *Excitation-Contraction Coupling and Cardiac Contractile Force* (Kluwer, Dordrecht, 2001).
- Marks, A. R., Priori, S., Memmi, M., Kontula, K. & Laitinen, P. J. Involvement of the cardiac ryanodine receptor/calcium release channel in catecholaminergic polymorphic ventricular tachycardia. *J. Cell Physiol.* **190**, 1–6 (2002).
- Blaustein, M. P. & Lederer, W. J. Sodium/calcium exchange: its physiological implications. *Physiol. Rev.* **79**, 763–854 (1999).
- Reuter, H. *et al.* The Na⁺-Ca²⁺ exchanger is essential for the action of cardiac glycosides. *Circ. Res.* **90**, 305–308 (2002).
- Philipson, K. D. & Nicoll, D. A. Sodium-calcium exchange: a molecular perspective. *Annu. Rev. Physiol.* **62**, 111–133 (2000).
- Santana, L. F., Kranias, E. G. & Lederer, W. J. Calcium sparks and excitation-contraction coupling in phospholamban-deficient mouse ventricular myocytes. *J. Physiol.* **503**, 21–29 (1997).
- Santana, L. F., Gomez, A. M. & Lederer, W. J. Ca²⁺ flux through promiscuous cardiac Na⁺ channels: slip-mode conductance. *Science* **279**, 1027–1033 (1998).
- Gomez, A. M. *et al.* Defective excitation-contraction coupling in experimental cardiac hypertrophy and heart failure. *Science* **276**, 800–806 (1997).
- duBell, W. H., Lederer, W. J. & Rogers, T. B. Dynamic modulation of excitation-contraction coupling by protein phosphatases in rat ventricular myocytes. *J. Physiol.* **493**, 793–800 (1996).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank the type 4 long-QT family for participation. Research support was provided by the Howard Hughes Medical Institute, NIH, the Muscular Dystrophy Association, Canadian Institutes of Health, the Institut National de la Santé et de la Recherche Médicale (INSERM), and the programme Hospitalier de Recherche Clinique. We also thank C. Kontos and B. Knollman for discussions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.B. (e-mail: benne012@mc.duke.edu).

NF- κ B blockade and oncogenic Ras trigger invasive human epidermal neoplasia

Maya Dajee*, Mirella Lazarov*, Jennifer Y. Zhang, Ti Cai, Cheryl L. Green, Alan J. Russell, M. Peter Marinkovich, Shiyong Tao, Qun Lin, Yoshiaki Kubo & Paul A. Khavari

Veterans Affairs Palo Alto Healthcare System and the Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California 94305, USA
* These authors contributed equally to this work

The nuclear factor NF- κ B and oncogenic Ras can alter proliferation in epidermis, the most common site of human cancer^{1,2}. These proteins are implicated in epidermal squamous cell carcinoma in mice^{3–5}, however, the potential effects of altering their function are uncertain. Whereas inhibition of NF- κ B enhances apoptosis in certain tumours⁶, blockade of NF- κ B predisposes murine skin to squamous cell carcinoma^{5,7}. Because therapeutics inhibiting Ras and NF- κ B pathways are being developed to treat human cancer^{8,9}, it is essential to assess the effects of altering these regulators. The medical relevance of murine studies is limited, however, by differences between mouse and human skin, and by the greater ease of transforming murine cells. Here we show that in normal human epidermal cells both NF- κ B and oncogenic Ras trigger cell-cycle arrest. Growth arrest triggered by oncogenic Ras can be bypassed by I κ B α -mediated blockade of NF- κ B, generating malignant human epidermal tissue resembling squamous cell carcinoma. Human cell tumorigenesis is dependent on laminin 5 and α 6 β 4 integrin. Thus, I κ B α circumvents restraints on growth promotion induced by oncogenic Ras and can act with Ras to induce invasive human tissue neoplasia.

To study Ras and NF- κ B in a setting more relevant to human tumorigenesis, we expressed the active Ha-Ras Gly12Val mutant, NF- κ B p65 and a stable NF- κ B repressor mutant of I κ B α ¹⁰ in human skin tissue. Primary human keratinocytes were retrovirally transduced and used to regenerate human skin on immune-deficient mice¹¹. Tissue expressing I κ B α alone showed mild hyperplasia, whereas expression of oncogenic Ras induced growth arrest with graft failure (Fig. 1a and Supplementary Table 1). Although implicated in promoting features of neoplasia in other settings¹², the coexpression of oncogenic Ras with NF- κ B subunits failed to support proliferation (Supplementary Fig. 1a). Notably, the coexpression of Ras and I κ B α (Ras-I κ B α) produced large neoplasms resembling human squamous cell carcinomas (SCCs) in 3 weeks (Fig. 1a and Supplementary Fig. 1b).

Rapidly growing Ras-I κ B α tumours showed hallmarks of SCC. Histologically, Ras-I κ B α human epidermis showed massive neoplasia with deep invasion through fat and into underlying muscle

and fascia (Fig. 1a). There are no diagnostic markers for human SCC²; however, Ras-I κ B α epidermal tumours showed changes in protein expression that have been reported as characteristic for SCC, including increases in vascular endothelial growth factor (VEGF) and matrix metalloproteinase 3 (MMP3), and a decrease in E-cadherin (Fig. 1b). Ras-I κ B α tumours also showed a more than tenfold increase in mitotic index (Fig. 1b). Thus, Ras-I κ B α epidermal neoplasia strongly resembles spontaneous human SCC.

To determine the relevance of these findings to epidermal cancer, we studied primary human cutaneous SCCs. p65 was redistributed to the cytoplasm both in Ras-I κ B α tumours and in SCCs obtained from patients (Fig. 1c). In addition, immunoblots from ten consecutive SCCs showed that a subset of SCCs expressed increased

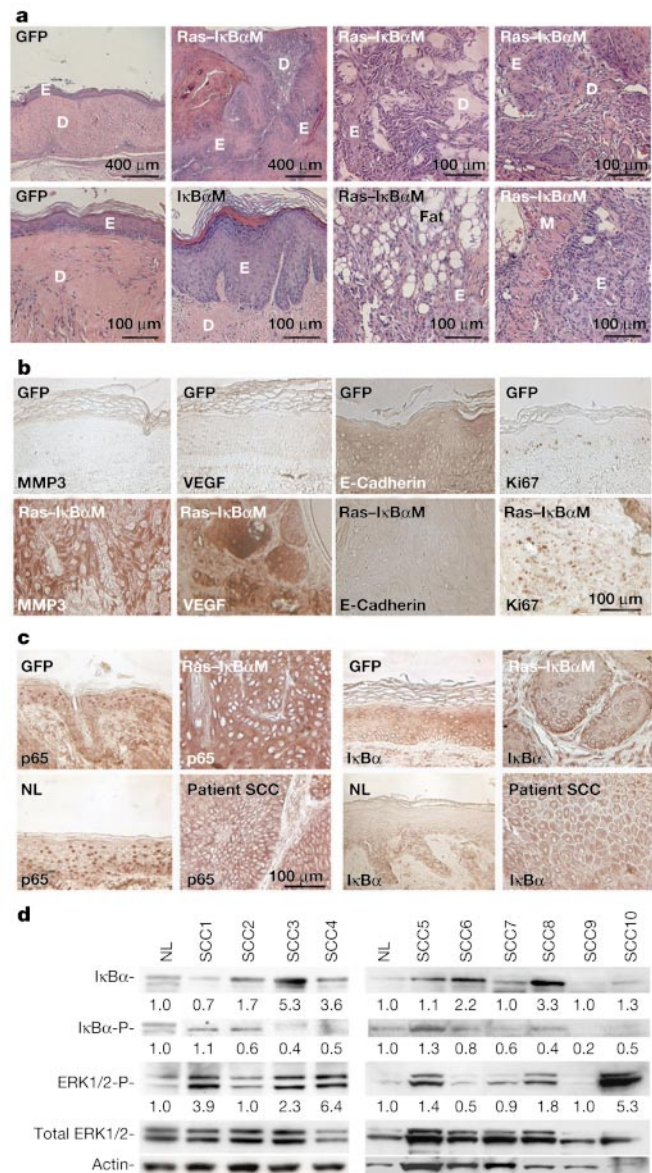


Figure 1 Ras and I κ B α human epidermal neoplasia. **a**, Histology. Invasive epidermal neoplasia (E) through dermis (D), fat and muscle (M). **b**, Expression of MMP3, VEGF, E-cadherin and Ki67. **c**, Expression of p65 and I κ B α in regenerated human epidermal tissue expressing either Ras-I κ B α or GFP as compared with normal patient skin (NL) and spontaneously arising SCC. Note the loss of nuclear p65 distribution in both Ras-I κ B α -induced and spontaneous epidermal tumours, and the widespread distribution of I κ B α . **d**, Immunoblots of ten additional consecutive primary cutaneous SCCs.

amounts of I κ B α (Fig. 1d), consistent with reports of NF- κ B inhibition in human epithelial breast cancers¹³. Increased expression and distribution of both Ras and its phosphorylated mitogen-activated protein kinase (MAPK) targets was seen in Ras-I κ B α tumours and patient SCCs (Supplementary Fig. 1d), and a subset of patient SCCs showed increased activation of the Ras target extracellular regulated signal kinases 1 and 2 (ERK1/2; Fig. 1d). Quantities of active Ras-GTP were comparable in Ras-I κ B α tumours and in human epithelial cancer cells with defined Ras mutations¹⁴, indicating that tumorigenesis was not due to expression of more active Ras than is expressed by endogenous mutant alleles. These findings support the potential relevance of both activation of the Ras/MAPK pathway and blockade of NF- κ B in at least a subset of spontaneous human epidermal SCCs.

To explore how Ras and I κ B α generate neoplasia, we studied processes that are important in tumorigenesis, including proliferation, apoptosis and matrix interactions. When coexpressed with Ras in primary human epidermal cells, I κ B α bypassed Ras-triggered G1 arrest (Fig. 2a, b). Ras itself induced several changes that promote G1 arrest, including induction of cyclin-dependent kinase (CDK) inhibitors (CKIs) p21^{CIP1}, p15^{INK4B} and p16^{INK4A} (Fig. 2c). In addition, Ras decreased the concentrations of CDK4. These changes were associated with increased hypophosphorylated retinoblastoma protein (Rb) and decreased cyclin E, both of which inhibit G1 progression (Fig. 2c). Notably, coexpression of I κ B α reversed these Ras effects (Fig. 2c).

Because I κ B α inhibited NF- κ B but could also exert other effects, we examined whether NF- κ B activity contributes to these G1 regulator changes. NF- κ B reporter gene activity was induced by

Ras and inhibited by I κ B α (Fig. 3a), indicating that Ras activates NF- κ B in keratinocytes. This raised the possibility that blockade of Ras-induced NF- κ B activity might inhibit the effects of NF- κ B on G1 regulators. NF- κ B causes keratinocyte growth arrest and, because other studies have shown the importance of CKIs in this process¹⁵, we examined the effects of NF- κ B on CDKs. Notably, whereas I κ B α induced CDK4, NF- κ B suppressed it (Fig. 3b). If this CDK4 suppression is important in epidermal growth arrest triggered by Ras and NF- κ B, then preventing it should bypass growth inhibition. In fact, expression of CDK4 bypassed the growth arrest induced by both Ras and NF- κ B (Fig. 3c, d, and Supplementary Fig. 3a), suggesting that I κ B α -maintained expression of CDK4 during strong Ras signalling is functionally important in G1 progression, consistent with recent findings¹⁶.

Because sustained growth after such G1 escape requires telomere maintenance, we examined telomere length in Ras-I κ B α neoplasia. Ras-I κ B α tumours preserved normal telomere length at 8 weeks (Fig. 3e) and increased the expression of hTERT protein as compared with normal control tissue (Supplementary Fig. 3b), suggesting that mechanisms supporting telomere maintenance accompany tumorigenesis.

In addition to overcoming growth restraints, tumour cells acquire resistance to apoptosis. NF- κ B blockade, however, promotes epidermal apoptosis⁵. Because Ras can act through both the Raf/MAPK and phosphatidylinositol-3-OH kinase (PI(3)K)/Akt pathways to oppose apoptosis^{17,18}, we tested whether Ras reverses apoptosis induced by NF- κ B blockade. Coexpression of oncogenic Ras reduced the enhanced apoptosis mediated by tumour-necrosis factor- α (TNF- α) observed in I κ B α -expressing cells (Fig. 3f), indicating that Ras opposes the increased susceptibility to apoptosis caused by NF- κ B blockade.

Matrix interactions represent another important factor in carcinogenesis. Ectopic subcutaneous tumour formation in mice often requires administration of matrigel or fibroblast-derived matrix^{16,19}; however, Ras-I κ B α tumorigenesis required neither. Rapidly growing tumours developed within 2 weeks after subcutaneous injection of Ras-I κ B α cells alone, (Fig. 4a, b), suggesting that these cells are competent to generate their own supporting matrix and prompting us to study the requirements for specific keratinocyte adhesion molecules in Ras-I κ B α neoplasia.

Integrin-matrix interactions strongly influence epithelial tumorigenesis^{20,21}. α 6 β 4 integrin and its ligand laminin 5 are upregulated in SCC of skin, colon, oesophagus and larynx, and higher expression is correlated with increased invasiveness^{22,23}. In addition, Ras induction increases α 6 β 4 integrin concentrations in epidermal cells²⁴. We therefore explored potential cooperative effects of laminin 5 and α 6 β 4 integrin on Ras-I κ B α neoplasia. Laminin 5 and α 6 β 4 integrin are expressed strongly throughout Ras-I κ B α tumours (Fig. 4c), and we used inhibitory antibodies to assess their functional importance. Antibodies specific for human laminin 5 and for α 6 β 4 integrin, but not control immunoglobulin- γ (IgG), inhibited tumorigenesis (Fig. 4d and Supplementary Fig. 3c).

The importance of laminin 5 and α 6 β 4 integrin was verified further by expressing Ras and I κ B α in primary human keratinocytes from individuals affected with epidermolysis bullosa carrying mutations in genes encoding these proteins. Unlike people affected with epidermolysis bullosa caused by a *COL7A1* mutant, these individuals do not develop aggressive SCCs²⁵. Immunoblotting confirmed successful Ras-I κ B α gene transfer in these cells (Supplementary Fig. 4a). Both laminin-5-negative and β 4-integrin-negative human keratinocytes failed to form tumours on coexpression of Ras and I κ B α , however, reintroduction of the wild-type *LAMB3* and *ITGB4* genes into these cells restored their tumour-forming capacity (Fig. 4e, f), suggesting that these proteins are required for human SCC tumorigenesis.

Given the aggressive subcutaneous growth of Ras-I κ B α neoplasms, we next examined their capacity to disseminate to viscera

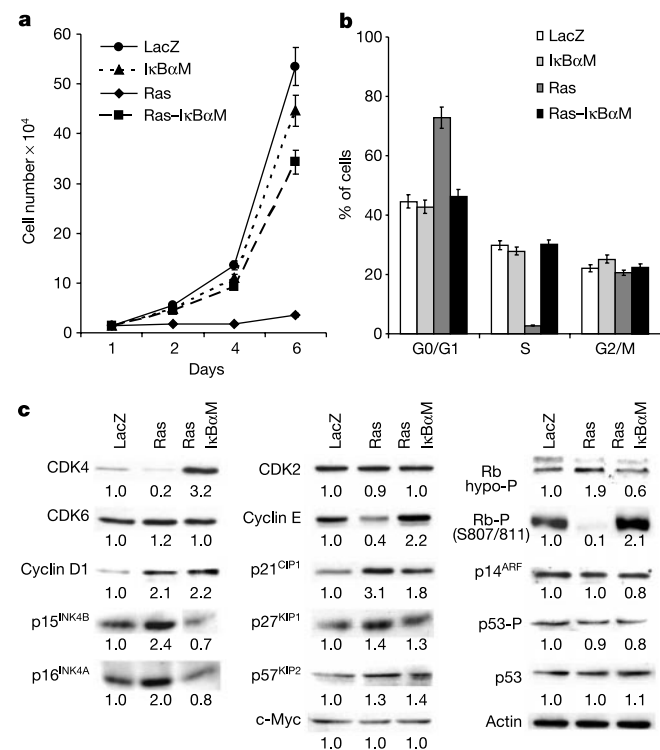


Figure 2 I κ B α rescues keratinocyte growth arrest triggered by oncogenic Ras. **a**, Proliferation of primary epidermal cells. Data are the means $\pm$ s.d. of triplicate independent transductions for each vector at each time point. **b**, Cell-cycle distribution 48 h after gene transfer as determined by flow cytometry. **c**, Expression of G1 regulator proteins as a function of oncogenic Ras and the impact of I κ B α coexpression. Cells were serially transduced with the vectors indicated at the top of the panels and immunoblotted with antibodies against the molecules indicated at the left of each panel.

when delivered to the bloodstream. Primary Ras-I κ B α keratinocytes generated several pulmonary nodules in all samples by 4 weeks (Fig. 4g). A wasting process resembling cachexia accompanied visceral tumorigenesis. Keratinocytes isolated from Ras-I κ B α tumours showed normal diploid karyotypes (Supplementary

Fig. 5), consistent with the premise that these tumours are driven by Ras-I κ B α rather than genetic instability, a possibility supported by a study confirming that the latter is not an invariant feature of human transformation²⁶.

I κ B α inhibits NF- κ B complexes containing p65 and c-Rel but

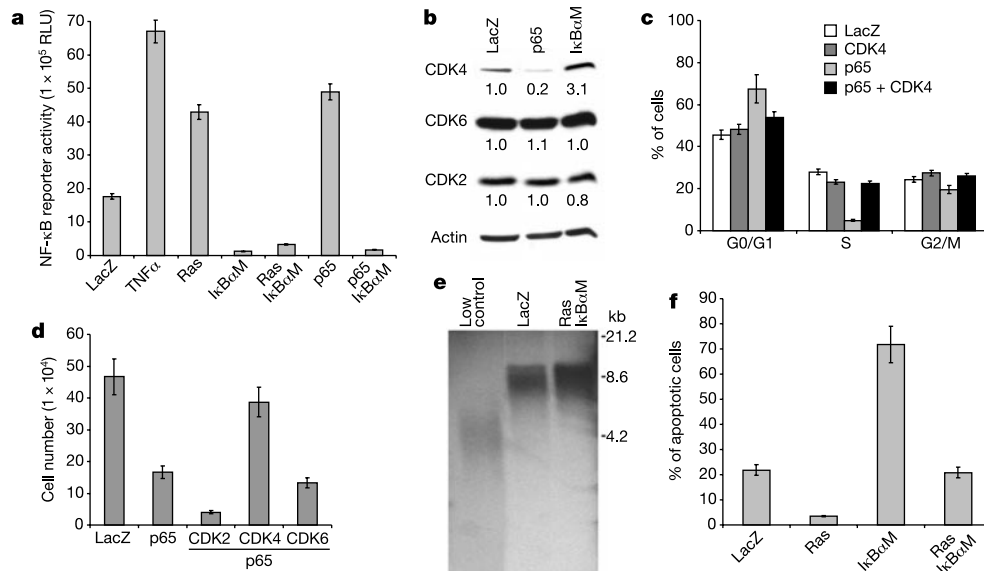


Figure 3 Effects of Ras, p65 and I κ B α on CDK expression, proliferation, telomere length and apoptosis. **a**, NF- κ B reporter activity. **b**, G1 CDK protein expression. **c**, Cell-cycle distribution 48 h after expression of p65, CDK4, p65 and CDK4 coexpression, and marker control. **d**, Cell numbers 72 h after retroviral transduction. **e**, Telomere southern blot of

LacZ-transduced control epidermis and Ras-I κ B α epidermal tumour tissue after 8 weeks *in vivo*; the low control represents genomic DNA from a cell line with short telomeres (mean length 3.9 kilobases). **f**, Susceptibility to TNF- α -mediated apoptosis as a function of Ras and I κ B α expression.

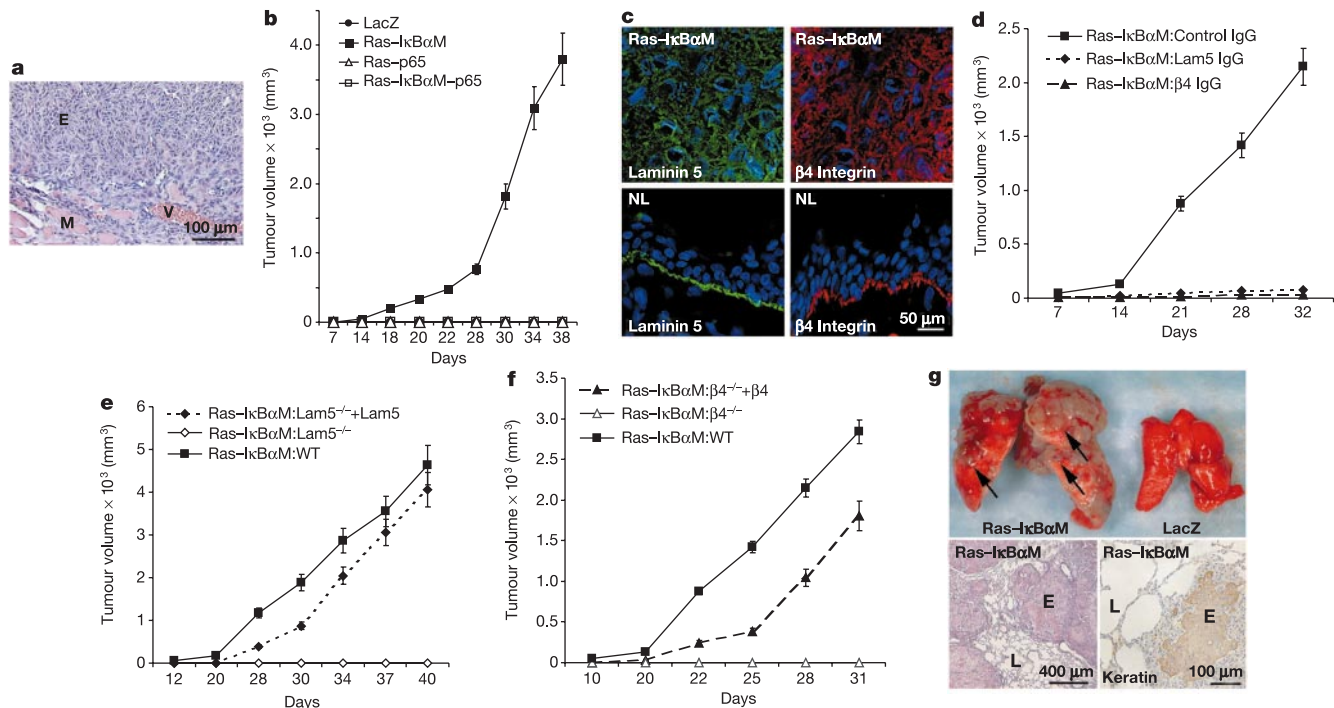


Figure 4 Ras-I κ B α tumorigenesis. **a**, Subcutaneous tumorigenesis. E, epidermal tissue; M, muscle; V, blood vessel. **b**, Tumour growth kinetics after subcutaneous injection of primary keratinocytes expressing the indicated proteins. **c**, Expression of laminin 5 (green) and α 6 β 4 integrin (red). Nuclei are stained with DAPI (4,6-diamidino-2-phenylindole dihydrochloride; blue). **d**, Laminin 5 and β 4 integrin blocking antibodies inhibit

tumorigenesis. **e**, **f**, Laminin-5-negative (**e**) and β 4-integrin-negative (**f**) primary human keratinocytes do not form tumours but regain tumorigenicity on protein restoration.

g, Primary keratinocytes expressing both Ras and I κ B α generate visceral tumours after intravenous injection. Top, lungs at 4 weeks; bottom left, histology (E, epidermal tissue; L, lung tissue); bottom right, antibodies specific for human keratin (brown).

may promote function of other NF- κ B dimers²⁷ and earlier work has suggested that p65 blockade and induction of nonclassical NF- κ B subunits may be important in breast carcinogenesis¹³. To examine whether I κ B α tumorigenesis could be explained by such a mechanism, we tested the tumorigenicity of Ras with all five NF- κ B subunits and found that expression of active nonclassical subunits with oncogenic Ras does not produce neoplasia (Supplementary Table II). I κ B α is best known for inhibiting specific NF- κ B subunits, however, it may function in other ways to promote tumorigenesis. To address this, we examined whether Ras-I κ B α tumour formation can be abrogated by active p65. We found that p65 blocks Ras-I κ B α tumorigenesis (Fig. 4b), indicating that NF- κ B inhibition, as opposed to other uncharacterized I κ B α functions, is the relevant mechanism of I κ B α -driven epidermal neoplasia.

Coexpression of Ras and I κ B α thus cooperate to induce human epidermal neoplasia. Our data do not establish Ras and I κ B α as inducers of most SCCs. Instead, these proteins may engage programmes of carcinogenesis that are accessible to several oncogenic factors. Although it mediates pleiotropic neoplastic effects^{24,28}, Ras by itself cannot override epidermal G1 growth restraints. I κ B α bypasses Ras G1 restraints, namely CKI upregulation and CDK4 repression, to permit unrestrained proliferation. In this process, Ras reduces susceptibility to apoptosis mediated by NF- κ B blockade and Ras-I κ B α tumours show preserved telomeres and increased amounts of hTERT protein. Although the mechanism responsible for the latter requires further study, Ras effects on factors implicated in transcription of hTERT²⁹ are of interest.

Our data suggest that interactions among α 6 β 4 integrin, Ras and NF- κ B can exert dominant effects on epidermal growth control. Taken together, these observations suggest that positive reciprocal interactions operate between α 6 β 4 integrin and Ras in keratinocytes and that blockade of Ras-induced NF- κ B in this setting is tumorigenic. The neoplastic effects of epidermal NF- κ B blockade underscore the pitfalls facing oncotherapeutics designed to enhance tumour apoptosis by inhibiting NF- κ B^{6,8}. □

Methods

Cell culture and gene transfer

Complementary DNAs encoding human Ha-Ras G12V, I κ B α M, p65, p50, p52, c-Rel, RelB, CDK2, CDK4, CDK6, laminin 5 and the β 4 integrin subunit, as well as p65-IRES-I κ B α M sequences for the simultaneous coexpression of Ras, I κ B α M and p65 were subcloned in the LZRS retroviral vector, and virus was produced as described¹¹. Primary human epidermal cells isolated from normal human skin were transduced with retroviral vectors with subsequent transductions repeated at intervals of 8–12 h. For epidermolysis bullosa cell studies, primary keratinocytes from genetically characterized individuals with mutations in *LAMB3* (encoding laminin 5 β 3) and *ITGB4* (encoding the β 4 integrin subunit) were used. We verified that gene transfer efficiency was higher than 99% by immunostaining as described¹¹. Protein expression driven by these vectors was further confirmed by immunoblotting in each experiment. For all studies, triplicate independent transductions were analysed, and data are expressed as means $\pm$ s.d. For tumour studies, cells were used from six unrelated donors with similar results for all samples. For proliferation studies, triplicate independent transductions of 1×10^4 cells per 35-mm plate were collected and counted at 24-h intervals.

Protein expression

We used antibodies specific for the following molecules: CDK4, CDK6, p65, p50, p52, c-Rel, RelB, cyclin A, p53, p18^{INK4C}, p19^{INK4D}, p21^{CIP1}, p57^{KIP2}, p14^{ARF}, CDC25A, Ha-Ras (Santa Cruz Biotechnology); I κ B α (Calbiochem), CDK2, cyclin D1, cyclin E, hypophosphorylated-Rb, c-Myc, p16^{INK4A} and p27^{KIP1} (PharMingen); p15^{INK4B}, VEGF-C (Upstate Biotechnologies); hTERT, MMP-1, MMP2 and MMP3 (Oncogene); Ki67 (Novocastra); E-cadherin (Zymed); ERK1/2, p53, phospho-Rb (Cell Signaling Technology); and β 4 integrin subunit (Chemicon). We also used antibodies against the β 3 chain of laminin 5 (K140) and BM165 blocking antibodies against laminin 5 (ref. 30). Sixty micrograms of whole-cell protein extract was loaded per lane for immunoblotting. A GS-710 Calibrated Imaging Densitometer (Bio-Rad) was used for optical densitometric quantification, and data were normalized to β -actin and total ERK controls obtained from blots that had been stripped and re-probed for the same sample lane. Patient tissue was obtained from freshly excised, pathologically confirmed SCC tissue specimens from 14 consecutive unrelated patients; 4 were analysed by immunohistochemistry and 10 were used to prepare protein extracts for immunoblotting. We verified antibody specificity for p65 using RelA-deficient tissue. TdT-mediated dUTP nick end labelling (TUNEL)-positive cells were scored with mean $\pm$ s.d. apoptotic indices as described¹⁶.

Animal studies

Genetically engineered human epidermis was regenerated on CB.17 *scid/scid* mice after retroviral gene transfer as described¹¹. We carried out grafts on 4–8 mice per vector group. At 4–8 weeks after grafting, human skin tissue was excised and subjected to analysis. Staining with antibodies specific for human keratin, as well as immunostaining for Ras and I κ B α , confirmed that all deeply invasive tissue was of human epidermal origin.

Tumorigenicity assays were done as described¹⁹; 8-week-old athymic nude mice (five per vector group) were γ -irradiated with 400 rad then injected 4 h later either subcutaneously or intravenously with 1×10^6 cells resuspended in 100 μ l of PBS. Initial subcutaneous tumour formation was defined as the time when the tumour had reached 3 mm in diameter. Tumour growth was followed for 6–8 weeks; before this time point was reached, however, some mice died from massive tumour burden or were killed for humane protocol requirements. For antibody blocking studies, after subcutaneous injection of cells expressing Ras and I κ B α , mice received three weekly, 0.5-mg injections of purified antibodies to either the β 4 integrin subunit or laminin 5. The presence of injected antibodies was confirmed on normal human skin with collected sera. Tumour nodules formed in all animals injected intravenously with primary epidermal cells expressing Ras-I κ B α , at an average number of 30.2 clinically visible tumour nodules per animal; tissue from mice injected with marker control cells showed no detectable abnormalities on either gross or microscopic examination.

Received 26 June; accepted 25 November 2002; doi:10.1038/nature01283.

- Miller, D. L. & Weinstock, M. A. Nonmelanoma skin cancer in the United States: incidence. *J. Am. Acad. Dermatol.* **30**, 774–778 (1994).
- Alam, M. & Ratner, D. Cutaneous squamous-cell carcinoma. *N. Engl. J. Med.* **344**, 975–983 (2001).
- Yuspa, S. H. The pathogenesis of squamous cell cancer: lessons learned from studies of skin carcinogenesis. The thirty-third G. H. A. Clowes Memorial Award Lecture. *Cancer Res.* **54**, 1178–1189 (1994).
- Pazzaglia, S. *et al.* Analysis of c-Ha-ras gene mutations in skin tumors induced in carcinogenesis-susceptible and carcinogenesis-resistant mice by different two-stage protocols or tumour promoter alone. *Mol. Carcinog.* **30**, 111–118 (2001).
- van Hogerlinden, M., Rozell, B. L., Ahrlund-Richter, L. & Toftgard, R. Squamous cell carcinomas and increased apoptosis in skin with inhibited Rel/nuclear factor- κ B signaling. *Cancer Res.* **59**, 3299–3303 (1999).
- Yamamoto, Y. & Gaynor, R. B. Therapeutic potential of inhibition of the NF- κ B pathway in the treatment of inflammation and cancer. *J. Clin. Invest.* **107**, 135–142 (2001).
- Seitz, C. S., Lin, Q., Deng, H. & Khavari, P. A. Alterations in NF- κ B function in transgenic epithelial tissue demonstrate a growth inhibitory role for NF- κ B. *Proc. Natl Acad. Sci. USA* **95**, 2307–2312 (1998).
- Karin, M., Cao, Y., Greten, F. R. & Li, Z. W. NF- κ B in cancer: from innocent bystander to major culprit. *Nature Rev. Cancer* **2**, 301–310 (2002).
- Adjei, A. A. Blocking oncogenic Ras signaling for cancer therapy. *J. Natl Cancer Inst.* **93**, 1062–1074 (2001).
- Van Antwerp, D. J., Martin, S. J., Kafri, T., Green, D. R. & Verma, I. M. Suppression of TNF- α -induced apoptosis by NF- κ B. *Science* **274**, 787–789 (1996).
- Robbins, P. B. *et al.* In vivo restoration of laminin 5 β 3 expression and function in junctional epidermolysis bullosa. *Proc. Natl Acad. Sci. USA* **98**, 5193–5198 (2001).
- Mayo, M. W. *et al.* Requirement of NF- κ B activation to suppress p53-independent apoptosis induced by oncogenic Ras. *Science* **278**, 1812–1815 (1997).
- Cogswell, P. C., Guttridge, D. C., Funkhouser, W. K. & Baldwin, A. S. Jr Selective activation of NF- κ B subunits in human breast cancer: potential roles for NF- κ B2/p52 and for Bcl-3. *Oncogene* **19**, 1123–1131 (2000).
- Mitchell, C. E., Belinsky, S. A. & Lechner, J. F. Detection and quantitation of mutant K-ras codon 12 restriction fragments by capillary electrophoresis. *Anal. Biochem.* **224**, 148–153 (1995).
- Seitz, C. S., Deng, H., Hinata, K., Lin, Q. & Khavari, P. A. Nuclear factor κ B subunits induce epithelial cell growth arrest. *Cancer Res.* **60**, 4085–4092 (2000).
- Lazarov, M. *et al.* CDK4 coexpression with Ras generates malignant human epidermal tumorigenesis. *Nature Med.* **8**, 1105–1114 (2002).
- Bonni, A. *et al.* Cell survival promoted by the Ras-MAPK signaling pathway by transcription-dependent and -independent mechanisms. *Science* **286**, 1358–1362 (1999).
- Stambolic, V., Mak, T. W. & Woodgett, J. R. Modulation of cellular apoptotic potential: contributions to oncogenesis. *Oncogene* **18**, 6094–6103 (1999).
- Elenbaas, B. *et al.* Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev.* **15**, 50–65 (2001).
- Watt, F. M. Role of integrins in regulating epidermal adhesion, growth and differentiation. *EMBO J.* **21**, 3919–3926 (2002).
- Mercuro, A. M. & Rabinovitz, I. Towards a mechanistic understanding of tumour invasion—lessons from the α 6 β 4 integrin. *Semin. Cancer Biol.* **11**, 129–141 (2001).
- Savoia, P., Trusolino, L., Pepino, E. & Marchisio, P. C. Expression and topography of integrins and basement membrane proteins in epidermal carcinomas: basal but not squamous cell carcinomas display loss of α 6 β 4 and BM-600/nicein. *J. Invest. Dermatol.* **101**, 352–358 (1993).
- Nordemar, S. *et al.* Laminin-5 as a predictor of invasiveness in cancer in situ lesions of the larynx. *Anticancer Res.* **21**, 509–512 (2001).
- Dajee, M., Tarutani, M., Deng, H., Cai, T. & Khavari, P. A. Epidermal Ras blockade demonstrates spatially localized Ras promotion of proliferation and inhibition of differentiation. *Oncogene* **21**, 1527–1538 (2002).
- Marinkovich, M. P., Herron, G. S., Khavari, P. A. & Bauer, E. A. *Dermatology in General Medicine* (ed. Freedberg, I. M.) 690–702 (McGraw-Hill, New York, 1999).
- Zimonjic, D., Brooks, M. W., Popescu, N., Weinberg, R. A. & Hahn, W. C. Derivation of human tumour cells in vitro without widespread genomic instability. *Cancer Res.* **61**, 8838–8844 (2001).
- Ghosh, S. & Karin, M. Missing pieces in the NF- κ B puzzle. *Cell* **109** (Suppl.), S81–S96 (2002).
- Pruitt, K. & Der, C. J. Ras and Rho regulation of the cell cycle and oncogenesis. *Cancer Lett.* **171**, 1–10 (2001).
- Wick, M., Zubov, D. & Hagen, G. Genomic organization and promoter characterization of the gene encoding the human telomerase reverse transcriptase (hTERT). *Gene* **232**, 97–106 (1999).

30. Marinkovich, M. P., Lunstrum, G. P. & Burgeson, R. E. The anchoring filament protein kalinin is synthesized and secreted as a high molecular weight precursor. *J. Biol. Chem.* **267**, 17900–17906 (1992).

Supplementary Information accompanies the paper on *Nature's* website (<http://www.nature.com/nature>).

Acknowledgements We thank H. Gladstone, M. Cooper, E. Fincher and Z. Cooper for SCC specimens; A. Anguiano for assistance with SKY-FISH; and N. Griffiths and P. Bernstein for administrative support. This work was supported by the US Veterans Affairs Office of Research and Development and by grants from the NIH to P.A.K. We acknowledge the generous support of the Epidermolysis Bullosa Research Foundation and the Nu Skin Center for Research at Stanford.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.A.K. (e-mail: khavari@CMGM.stanford.edu).

Telomere dysfunction and *Atm* deficiency compromises organ homeostasis and accelerates ageing

Kwok-Kin Wong*, Richard S. Maser*†, Robert M. Bachoo*‡, Jayant Menon§, Daniel R. Carrasco*||, Yansong Gu¶, Frederick W. Alt#☆ & Ronald A. DePinho*†

* Department of Adult Oncology, Dana Farber Cancer Institute, † Department of Genetics and Medicine, Harvard Medical School, ‡ Department of Neurology, Brigham and Women's Hospital, || Department of Pathology, Brigham and Women's Hospital, # The Center for Blood Research, Harvard Medical School, and ☆ Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA

§ University of California Los Angeles Medical School, Los Angeles, California 92507, USA

¶ Department of Radiation Oncology and Immunology, University of Washington School of Medicine, Seattle, Washington 98195, USA

Ataxia-telangiectasia (A-T) results from the loss of ataxia-telangiectasia mutated (*Atm*) function and is characterized by accelerated telomere loss, genomic instability, progressive neurological degeneration, premature ageing and increased neoplasia incidence¹. Here we evaluate the functional interaction of *Atm* and telomeres *in vivo*. We examined the impact of *Atm* deficiency as a function of progressive telomere attrition at both the cellular and whole-organism level in mice doubly null for *Atm* and the telomerase RNA component (*Terc*)^{2–4}. These compound mutants showed increased telomere erosion and genomic instability, yet they experienced a substantial elimination of T-cell lymphomas associated with *Atm* deficiency. A generalized proliferation defect was evident in all cell types and tissues examined, and this defect extended to tissue stem/progenitor cell compartments, thereby providing a basis for progressive multi-organ system compromise, accelerated ageing and premature death. We show that *Atm* deficiency and telomere dysfunction act together to impair cellular and whole-organism viability, thus supporting the view that aspects of A-T pathophysiology are linked to the functional state of telomeres and its adverse effects on stem/progenitor cell reserves.

Atm has a central role in signalling the presence of DNA double-strand breaks, and in the coordination of the appropriate checkpoint response functions linked in part to ATM-directed phosphorylation/activation of p53 and a host of cellular DNA-damage-response proteins⁵. *Atm* also participates in telomere maintenance, as demonstrated by accelerated cell-cycle-dependent telomere loss

and increased chromosomal end-to-end fusions in *Atm*-deficient mice and human cells^{6,7}. Human cell-culture-based studies have uncovered ATM- and p53-dependent checkpoint responses elicited by compromised telomere structure and function, brought about by a trans-dominant negative form of TRF2 (ref. 8). Several *Atm* mutant mouse strains have been generated; however, relative to A-T patients, a thorough phenotypic analysis of these mutant mice has revealed only modest neuronal and premature ageing sequelae, including a modest growth retardation and subtle motor coordination learning defects^{4,9–11}. The muted phenotypes have prompted speculation of cross-species differences, such as functional redundancy among members of the phosphatidylinositol-3-OH kinase (PI(3)K) family and/or increased telomere reserves in mice¹².

To dissect genetically the extent to which limiting telomeres are

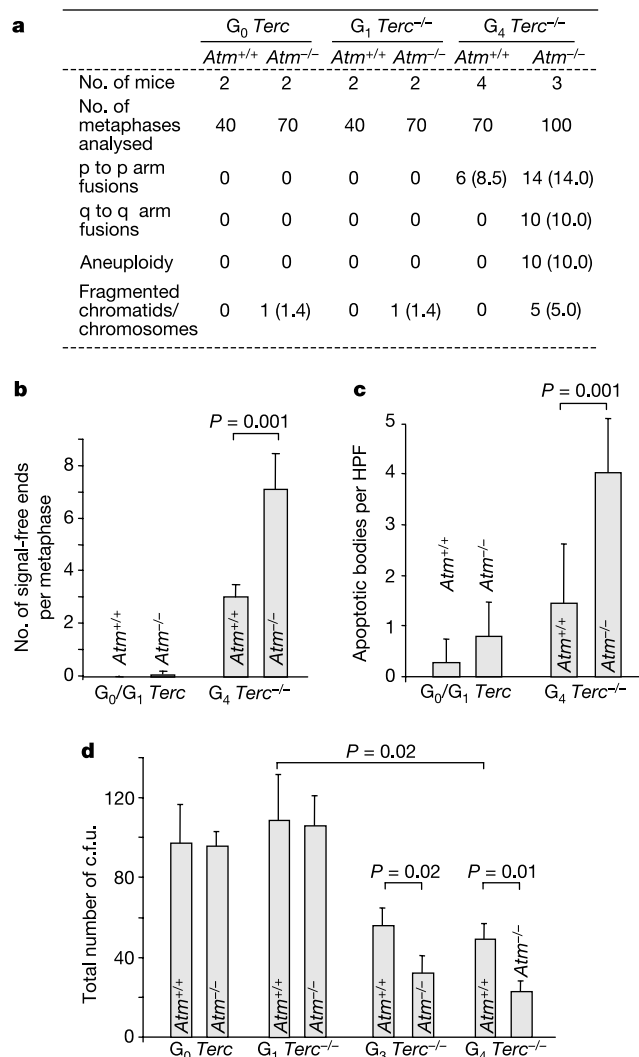


Figure 1 Increased chromosomal aberrations and signal-free ends in chromosome spreads from primary bone marrow cultures of late-generation *Atm*^{-/-} mice with resulting increased apoptosis and decreased proliferative potential. The experiments were performed with at least two mice per genotype. **a**, Tabular compilation of cytogenetic abnormalities. Numbers in parentheses are percentages of total metaphases. **b–d**, Histogram representations of the number of signal-free ends in the metaphases from the various mice cohorts (**b**), the number of apoptotic bodies per high-power field in the bone marrow of mice from the various cohorts (**c**), and the total number of colony formation units (c.f.u.) formed *in vitro* from bone marrow cell suspensions (**d**). G₀ is *Terc*^{-/-} *+/+* and G₁ is *Terc*^{-/-}.

responsible for or modulate the impact of *Atm* deficiency *in vivo*, we generated a series of cohorts with varying telomere length and function by breeding the mutant *Atm* allele⁴ through successive generations of *Terc*^{-/-} mice (Supplementary Fig. 1)^{2,13}. Metaphase spreads of primary bone marrow cultures derived from *Terc*^{+/-} or *Terc*^{-/-} (designated G₀) mice and from first generation (G₁) *Terc*^{-/-} mice, that were either wild-type or null for *Atm*, were diploid and showed minimal chromosome structural abnormalities (Fig. 1a). In contrast, fourth generation (G₄) *Terc*^{-/-} *Atm*^{-/-} metaphase spreads showed a significant increase in the number of chromosomal fusions and fragmented chromatids/chromosomes as compared with littermate G₄ *Terc*^{-/-} *Atm*^{+/-} samples (Fig. 1a; see also Supplementary Fig. 2a). In particular, G₄ *Terc*^{-/-} *Atm*^{-/-} metaphase spreads possessed abundant q-q arm (long arm to long arm) fusions as opposed to the near exclusive occurrence of p-p arm (short arm to short arm) (Robertsonian) fusions in G₄ *Terc*^{-/-} *Atm*^{+/-} metaphase spreads (Fig. 1a; see also Supplementary Fig. 2a)^{2,3,14}. Consistent with previous reports^{6,15,16}, telomere fluorescence *in situ* hybridization documented increased telomere signal-free ends in G₄ *Terc*^{-/-} *Atm*^{-/-} metaphase spreads compared with G₄ *Terc*^{-/-} *Atm*^{+/-} controls (7.2 compared with 3.0 per metaphase, respectively; *P* = 0.001, Fig. 1b).

The biological impact of the increased genomic instability stemming from *Atm* deficiency and telomere dysfunction was examined in bone marrow cells both *in vivo* and *in vitro*. In the bone marrow of G₄ *Terc*^{-/-} *Atm*^{-/-} mice, TdT-mediated dUTP nick end labelling and histological analyses showed a substantial increase in the number of apoptotic bodies relative to controls (Fig. 1c; see also Supplementary Fig. 2b, and data not shown). As reported previously², G₃/G₄ *Terc*^{-/-} *Atm*^{+/-} samples showed a moderate decline in the number and growth of all colony types in the haematopoietic colony-forming unit (c.f.u.) assays compared with the controls, and this decline was more pronounced in matched G₃/G₄ *Terc*^{-/-} *Atm*^{-/-} cultures (Fig. 1d). A cooperative impact was also evident on 3T9 passage of G₄ *Terc*^{-/-} *Atm*^{-/-} mouse

embryonic fibroblasts (MEFs), which barely attained two population doublings and rapidly senesced relative to controls (Supplementary Fig. 3a, b). Together, these findings establish that *Atm* deficiency and telomere dysfunction act cooperatively to limit the growth of distinct primary cell types and possibly haematopoietic progenitors. This finding prompted us to conduct a more detailed analysis of stem compartments in other organ systems.

The intestinal epithelium provides a highly quantitative system in which to assess growth and survival dynamics of a stem-cell compartment *in situ*¹⁷. Crypt apoptotic bodies were rare in G₀ and G₁ *Terc*^{-/-} *Atm*^{+/-} and *Atm*^{-/-} intestines but increased progressively in each successive generation of age-matched G₂ to G₄ *Terc*^{-/-} *Atm*^{+/-} mice (Fig. 2a) and with advancing age in G₃ *Terc*^{-/-} *Atm*^{+/-} mice (Fig. 2b). The increase in apoptosis through progressive *Terc* generations tracked closely with increasing telomere dysfunction as measured by anaphase bridging (Fig. 2c). The modest level of apoptosis in crypts of G₂ and G₃ *Terc*^{-/-} *Atm*^{+/-} mice was decreased in G₂ and G₃ *Terc*^{-/-} *Atm*^{-/-} controls, consistent with a role for *Atm* in the response to low levels of telomere dysfunction (Fig. 2a, c). In contrast, the higher levels of telomere dysfunction in G₄ *Terc*^{-/-} *Atm*^{+/-} and *Atm*^{-/-} samples were correlated with comparably high rates of apoptosis (Fig. 2a). The sustained apoptosis in crypts of G₄ *Terc*^{-/-} *Atm*^{-/-} mice contrasts significantly with the markedly attenuated telomere checkpoint response observed in crypts of *Terc*^{-/-} *p53*^{-/-} mice (and other systems) possessing a comparable number of anaphase bridges (Fig. 2a)¹³. Consistent with this robust *Atm*-independent response to telomere dysfunction, we observed comparably intense anti-p53 immunoreactivity in G₄ *Terc*^{-/-} *Atm*^{+/-} and *Terc*^{-/-} *Atm*^{-/-} crypt cells (Fig. 2d). These data, as well as the above haematopoietic and MEF profiles, demonstrate significant differences in the *Atm*- and *p53*-directed responses to telomere dysfunction *in vivo*.

The hallmark progressive neuronal phenotype of A-T patients and abnormal neurological features of G₃/G₄ *Terc*^{-/-} *Atm*^{-/-} mice (see below) prompted a quantitative study of the constitutively

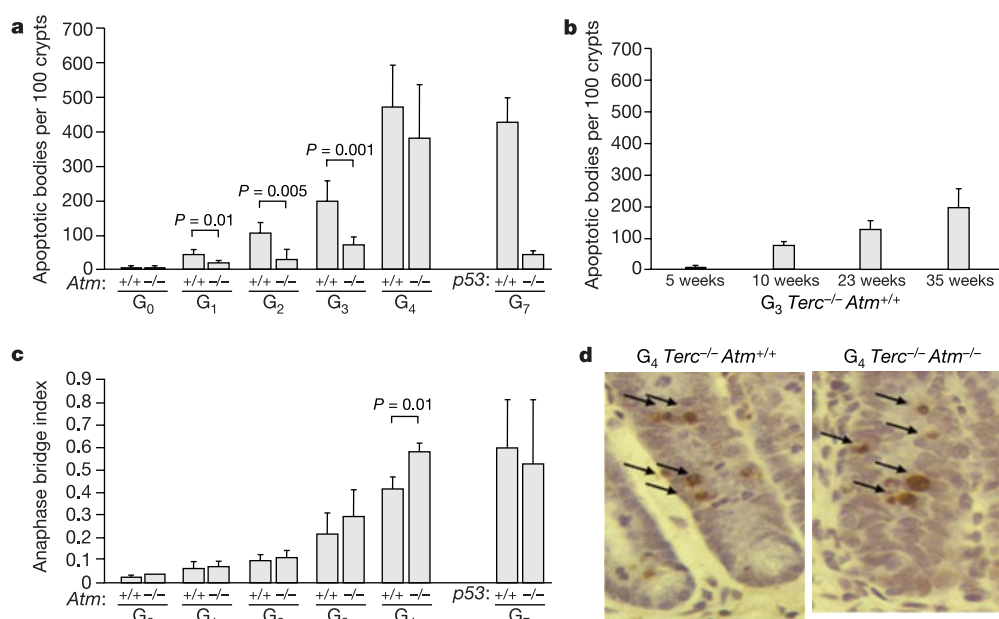


Figure 2 Increased apoptosis, anaphase bridging and p53 activation in the gastrointestinal crypts of G₃ and G₄ *Terc*^{-/-} *Atm*^{-/-} mice. Each of the experiments was performed with at least three mice per genotype. **a**, Number of apoptotic cells in the crypts from different mouse cohorts. **b**, Number of apoptotic cells in the crypts from G₃ *Terc*^{-/-} *Atm*^{+/-}

Atm^{+/-} mice of different ages. **c**, Per cent of metaphases with anaphase bridging in the crypts from different mice cohorts. **d**, Representative p53-positive staining in the gastrointestinal crypts of the G₄ *Terc*^{-/-} *Atm*^{+/-} and G₄ *Terc*^{-/-} *Atm*^{-/-} mice. Black arrows indicate areas of p53-positive staining.

proliferating neural stem-cell compartment in the adult brain. Quantification of 5-bromodeoxyuridine (BrdU) immunoreactive cells through the forebrain subependymal region showed fewer proliferating cells in G_4 $Terc^{-/-}$ $Atm^{+/+}$ brains when compared with G_0 to G_1 $Terc^{-/-}$ $Atm^{+/+}$ brains ($P = 0.001$), and this decline was more pronounced in G_4 $Terc^{-/-}$ $Atm^{-/-}$ samples ($P = 0.01$; Fig. 3a). These *in vivo* proliferation patterns mirrored the capacity of adult subependymal neuronal stem cells of the various genotypes to generate classical neurospheres *in vitro* (Fig. 3b, e). Individual G_0 to G_1 $Terc^{-/-}$ $Atm^{-/-}$ neurospheres were capable of generating new neurospheres¹⁸ over four passages with no indication of abatement. In contrast, only 36% of G_4 $Terc^{-/-}$ $Atm^{-/-}$ neurospheres generated secondary neurospheres and virtually none were produced at passage three.

The pronounced impact of A-T on neuronal lineages prompted us to assess the ability of individual neurospheres to differentiate into neurons compared with astrocytes. In matrigel containing

brain-derived neurotrophic factor (BDNF; neurons) or 1% FBS (astrocytes)¹⁹, G_0/G_1 $Terc^{-/-}$ $Atm^{+/+}$ and $Atm^{-/-}$ neurospheres showed robust and normal patterns of neuronal (TUJ1⁺) and astrocytic (glial fibrillary acidic protein (GFAP⁺)) differentiation (Fig. 3c, e), whereas G_3/G_4 $Terc^{-/-}$ $Atm^{-/-}$ neurospheres generated only GFAP-positive cells (Fig. 3c, e). Together with the observed reduction in forebrain subependymal proliferation, these data point to a marked decline in neural stem cells and a selective impairment in neuronal differentiation and/or survival (see below).

A detailed examination of the brain cytoarchitecture demonstrated no significant degenerative changes spanning an age range of 11–38 weeks. When subjected to an array of functional tests to evaluate individual motoric functions, there was no difference among the various cohorts²⁰. However, an open field test appeared to show a greater delay in initiation of exploration (ambulation) out of an outlined square for the G_4 $Terc^{-/-}$ $Atm^{-/-}$ mice compared with G_0 $Terc^{+/+}$ $Atm^{-/-}$ controls. These hypokinetic

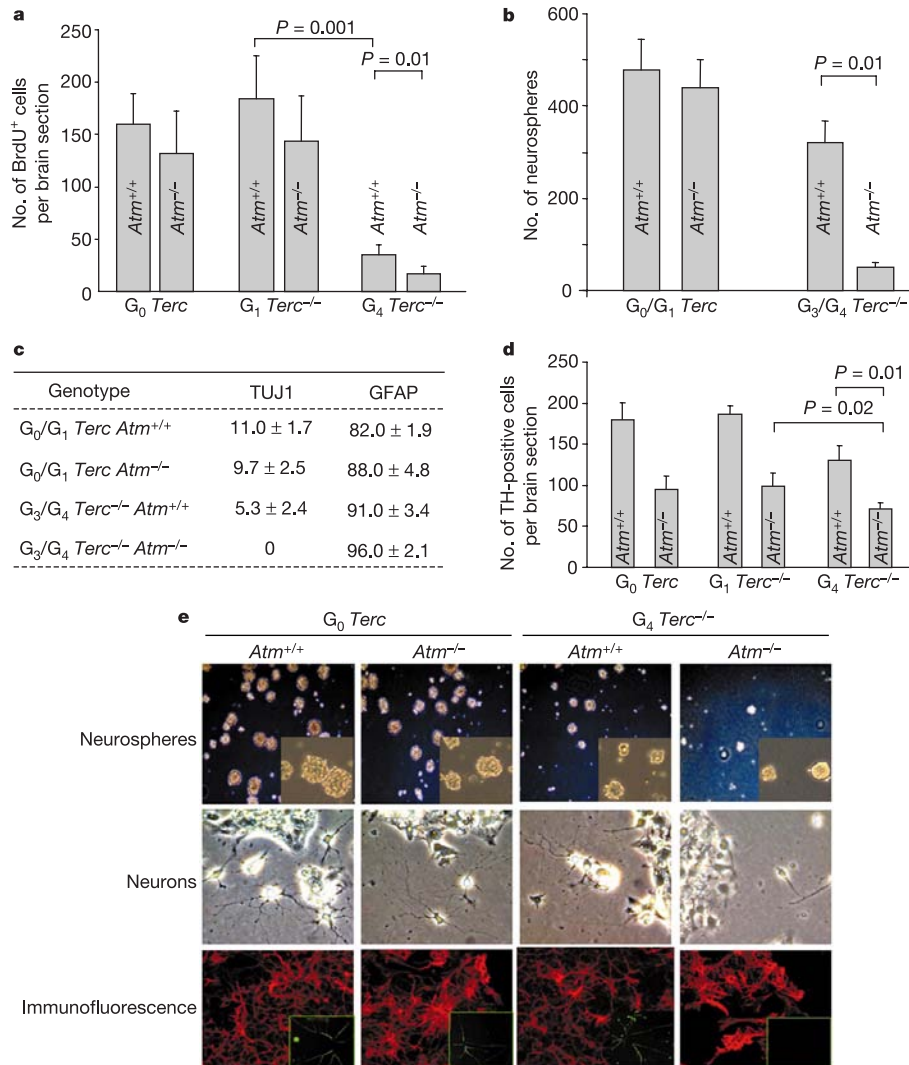


Figure 3 Influence of telomere function and *Atm* status on the proliferation and differentiation potential of the neural stem-cell compartment of adult brain. Each of the experiments was performed with at least three mice per genotype. G_0 is $Terc^{-/+}$ $+/+$ and G_1 is $Terc^{-/-}$. **a**, Number of cells staining positive for 5-bromodeoxyuridine (BrdU) in the subependymal zone of adult mice. **b**, Number of neurospheres generated *in vitro* from adult subependymal zone. **c**, Percentage of neurons (TUJ1 immunoreactive) and astrocytes (glial fibrillary acidic protein (GFAP) immunoreactive) generated from neural

stem-cell differentiation *in vitro*. **d**, Dopamine-containing neurons present in the substantia nigra pars compacta as measured by tyrosine hydroxylase (TH) immunoreactivity. **e**, Representative photographs of (top) neurospheres at $\times 10$ magnification (insets at $\times 20$ magnification), (middle) neurons ($\times 40$ magnification) and (bottom) GFAP-positive astrocytes ($\times 20$ magnification) with insets of TUJ1-positive neurons ($\times 40$ magnification), derived from neural stem cells isolated from various mouse cohorts.

movements observed in the late-generation *Terc*^{-/-} *Atm*^{-/-} mice suggested extrapyramidal tract dysfunction. Indeed, immunostaining for tyrosine hydroxylase revealed fewer dopamine-containing neurons in the substantia nigra pars compacta (SNc) and the ventral tegmental area (VTA) in *Atm*^{-/-} mice with intact telomeres²¹, and this decline was exacerbated further in *G*₄ *Terc*^{-/-} *Atm*^{-/-} mice (Fig. 3d). Hypokinetic movement disorder is of particular relevance, as extrapyramidal signs and cerebellar ataxia represent classical symptoms of A-T patients²².

The findings of diminished precursor/stem-cell reserve and increased apoptosis in several organ systems of late-generation *Terc*^{-/-} *Atm*^{-/-} mice prompted us to conduct a range of clinical and pathological studies. Hair greying and alopecia was increased in the *G*₄ *Terc*^{-/-} *Atm*^{-/-} cohorts as compared with the *G*₄ *Terc*^{-/-} *Atm*^{+/+} cohorts (45% compared with 30% of the total mice examined ($P = 0.001$), respectively; Supplementary Fig. 4 and Table 1). Similarly, there was a delay (three day difference) in hair re-growth, coagulum formation and complete healing in the *G*₄ *Terc*^{-/-} *Atm*^{+/+} and *G*₄ *Terc*^{-/-} *Atm*^{-/-} cohorts as compared to the corresponding *G*₀/*G*₁ cohorts. Lordokyphosis (hunchbacked spine) was detected at six months of age (three months earlier), and incidence was threefold higher at nine months of age (15% compared with 5%; $P = 0.001$) in *G*₄ *Terc*^{-/-} *Atm*^{-/-} mice relative to *G*₄ *Terc*^{-/-} *Atm*^{+/+} mice (Supplementary Table 1).

Weekly measurements of body weight revealed only modest loss

of weight for *G*₀/*G*₁ *Atm*^{-/-} and *G*₄ *Terc*^{-/-} mice compared with age- and gender-matched wild-type controls, as shown previously^{9,23}; however, marked reduction was observed in *G*₃/*G*₄ *Terc*^{-/-} *Atm*^{-/-} mice, which became more pronounced with advancing age (Supplementary Fig. 5a). The ratio of size and weight of internal organs of the various cohorts remained relatively constant throughout life, with the exception of aged *G*₃/*G*₄ *Terc*^{-/-} *Atm*^{-/-} mice possessing proportionally less muscle and fat mass (Supplementary Table 1). The overall histological architecture of the muscle fibres was found to be similar in the various cohorts. In the inguinal fat pads, although the *G*₃ to *G*₄ *Terc*^{-/-} *Atm*^{-/-} mice possessed a similar number of fat cells to the other cohorts, the mean fat cell size was found to be considerably smaller (Supplementary Fig. 4b). Notably, serial serum chemistry profiles and glucose tolerance tests were normal for all genotypes up to 38 weeks of age (Supplementary Table 1).

Overall survival analyses highlighted the complex interaction between *Atm* and telomere status. Consistent with a previous report⁴, *G*₀ *Terc*^{+/+} *Atm*^{-/-} and *G*₁ *Terc*^{-/-} *Atm*^{-/-} cohorts succumbed to thymic lymphoma with a median latency of 48 weeks (Fig. 4a), indicating that telomerase activity *per se* does not influence the tumour kinetics of *Atm* deficiency. In contrast, the presence of telomere dysfunction was associated with a near complete suppression of thymic lymphomas in *G*₃/*G*₄ *Terc*^{-/-} *Atm*^{-/-} mice. Despite cancer resistance, the median lifespan of the *G*₃/*G*₄ *Terc*^{-/-} *Atm*^{-/-} mice was significantly shorter than that of *G*₃/*G*₄ *Terc*^{-/-} *Atm*^{+/+} controls (Fig. 4b, c; see also Supplementary Table 1). Specifically, there was no discernible prodrome of accelerated weight loss, cachexia, reduced food and fluid intake or activity level. As is often the case with death in association with advanced age in humans, extensive analysis at autopsy failed to reveal a definitive cause of death (including cancer) in all *G*₃/*G*₄ *Terc*^{-/-} *Atm*^{-/-} mice.

This study reveals the complexity of the genetic interactions between *Atm* deficiency and telomere dysfunction *in vivo*. The marked differences between *Atm* and p53 in the setting of moderate to severe telomere dysfunction are made evident here by the contrasting clinical impact and cancer incidence between *Atm* and p53 mice with telomere dysfunction¹³. These studies also highlight the fact that the actions of *Atm* extend beyond its link to the p53 pathway, and that p53 can be activated independently of *Atm* by means of collateral pathways. In addition to activation of p53, *Atm* deficiency and telomere dysfunction cooperate to increase chromosomal instability, which may further compromise stem/progenitor cell function *in vivo*. Of note, epithelial cancers were not observed in the late-generation *Terc*^{-/-} *Atm*^{-/-} mutant mice in contrast to late-generation *Terc*^{-/-} p53 mutant mice²⁴—perhaps reflecting retention of an *Atm*-independent p53 checkpoint response to telomere dysfunction in those epithelial tissues. The *Atm*-independent p53 activation and apoptosis in stem/progenitor cell compartments, coupled with accelerated ageing and resistance to lymphoma, are reminiscent of the premature ageing and tumour resistance of mice harbouring an activated p53 germline allele²⁵. It is worth noting that the failure of *Atm* to attenuate the adverse cellular consequences of telomere dysfunction *in vivo* stands in contrast to previous observations in cultured human Epstein–Barr virus-transformed A-T B cells in which signalling pathways directed by telomere dysfunction involve the interaction of ATM with p53 (ref. 8). The basis for this difference is unclear and may relate to cross-species differences in *Atm* signalling from telomeres, *in vivo* versus *in vitro* systems, and/or differences in the signals elicited by telomere erosion versus altered telomere structure brought about by dominant-negative mutants of TRF2. We favour the latter explanation in light of the phenotypic differences brought about by loss of *Arf* function in the setting of telomere erosion²⁶ versus telomere structural alterations²⁷ in mouse cells. Finally, the disparate effects of *Atm* or p53 deficiency in the setting of telomere dysfunction seen

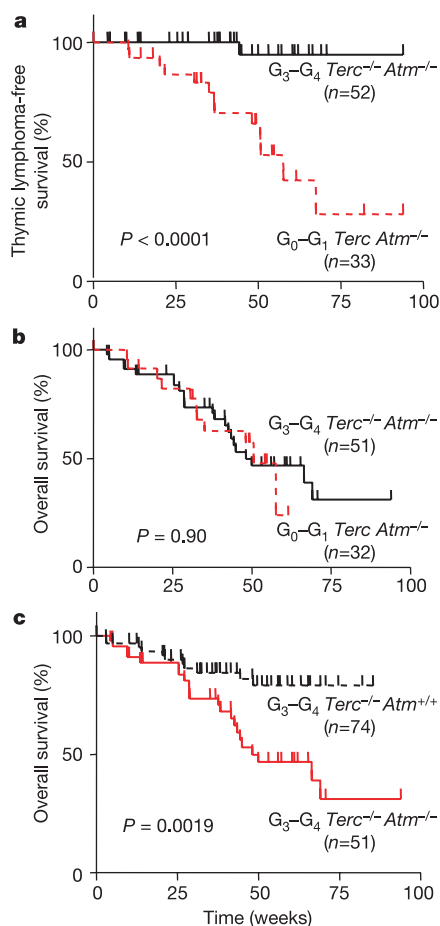


Figure 4 Decreased overall survival in late-generation *G*₃ to *G*₄ *Atm*^{-/-} mice. **a**, Kaplan–Meier curves of thymic lymphoma-free survival of *G*₀ to *G*₁ *Atm*^{-/-} and *G*₃ to *G*₄ *Atm*^{-/-} mice. **b**, Kaplan–Meier curves of the overall survival of *G*₀ to *G*₁ *Atm*^{-/-} and *G*₃ to *G*₄ *Atm*^{-/-} mice. **c**, Kaplan–Meier curves of the overall survival of *G*₃ to *G*₄ *Atm*^{+/+} and *G*₃ to *G*₄ *Atm*^{-/-} mice. *G*₀ is *Terc*^{+/+}, *+/+* and *G*₁ is *Terc*^{-/-}.

here and in earlier studies¹³ bears some similarities to the phenotypes of DNA ligase IV-deficient (*Lig4*^{-/-}) *Atm*^{-/-} and *Lig4*^{-/-} *p53*^{-/-} mice and cells^{28,29}. In this regard, telomere dysfunction can also cause impaired DNA repair³⁰.

With regard to A-T, our studies in mice provide possible insight into an important component of the neurodegenerative condition that afflicts patients. We show diminished capacity of late-generation *Terc*^{-/-} *Atm*^{-/-} neural stem cells to proliferate and differentiate into viable neurons as well as increased loss of dopamine-containing neurons in the late-generation compound mutant mice, establishing that telomere maintenance is essential to preserve neural stem/progenitor cell reserves and function. Of note, the classical A-T phenotype of ataxia was not found, raising the possibility that the cerebellar defects in A-T patients might be independent of the role of *Atm* in telomere dynamics, consistent with the clinical presentation of ataxia that occurs very early in life. It is tempting to speculate that the Purkinje cell defect relates to the other pleiotropic nuclear and cytoplasmic functions of *Atm*¹ and that these other activities may be shared by other members of the PI(3)K family of proteins in mice, but not in humans. Notwithstanding these differences, this model recapitulates several global aspects of A-T encountered in humans, such as multi-system failure and accelerated ageing. Moreover, a possible pathogenetic basis for these progressive multi-system defects is established by the correlation between telomere dysfunction and impaired stem-cell reserve and function. This model system affords an opportunity to dissect the complex pathogenetic contributions of *Atm* deficiency and telomeres in relation to various stem-cell compartments as well as providing a conceptual framework for conducting corresponding analyses in human A-T stem-cell compartments. It is clear that there remain important differences between this mouse model and the human condition; however, these differences will no doubt provide an entry point for a more precise dissection of the complexity of the *Atm*-p53-telomere axis in stem-cell renewal, ageing and cancer. □

Methods

Production of *Terc*^{-/-} *Atm* mice

The starting generation of *Terc*^{+/-} *Atm*^{+/-} (*G*₀) mice used in this study was generated from a cross between *Terc*^{+/-} *Atm*^{+/-} mice and third generation (*G*₃) *Terc*^{-/-} mice to reduce the number of generations to achieve critical telomere shortening in subsequent matings. These *G*₃ *Terc*^{-/-} mice did not show any cellular or cytogenetic signs of telomere dysfunction, indicating ample telomere reserve at the time of this initial cross. These mice were intercrossed to generate *G*₁ *Terc*^{-/-} *Atm*^{+/-}, *Atm*^{+/-} and *Atm*^{-/-} animals. All mice are in mixed genetic background WW6/C57BL/6, and cousin-mating schemes were used to prevent generation of substrains. *G*₁ *Terc*^{-/-} *Atm*^{+/-} mice were mated to generate *G*₂ mice, and so on.

Cytogenetic analysis

Metaphase spreads of primary bone marrow culture cells were prepared and processed by standard methods after activation by Methocult CF M3434 (Stem cell Technology) and colcemid treatment. For telomere PNA fluorescence *in situ* hybridization analysis, hybridization of Cy3-conjugated (TTAGGG)₃ (Perseptive Biosystems) was performed as described previously. We quantified structural chromosomal abnormalities on a Nikon Eclipse microscope equipped with a Spot camera and × 63 and × 100 objectives. Metaphases were scored with the experimenter blind to the genotype. We analysed at least 20 metaphases per genotype.

Mouse embryonic fibroblast assays

Mouse embryonic fibroblasts (MEFs) were prepared and 3T9 assays were carried out as previously described³.

Animal growth and tumour incidence

Mice from the various cohorts were weighed weekly until one year of age. Mice were examined closely for evidence of ill health or overt tumour growth. Mice were killed if profoundly ill or if external tumours exceeded 2 cm in diameter, and scored as a death in survival analysis. Only those animals with histologically proved cancer were scored as an event in the tumour incidence Kaplan-Meier analysis. Statistical significance was measured using the log-rank test. All mice were subjected to extensive autopsy and gross examinations. Tumours and various organs were fixed in 10% formalin and embedded in paraffin. Five-micrometre paraffin sections were stained with haematoxylin and eosin.

Alopecia, hair greying and wound-healing analysis

Hair greying and alopecia were quantified at different parts of the animal's body by estimating the percentage of grey hair or hair loss, and the results were normalized against the age-matched *G*₀ *Terc*^{+/-} *Atm*^{+/-} mice. Wound-healing studies were carried out by performing four 3-mm full-thickness punch biopsies extending through the epidermis and dermis to the panniculus carnosus using a circular skin punch. We measured wound sizes 2, 4, 6 and 8 days after wounding.

Haematopoietic colony assay

Murine colony-forming unit assay was performed according to the manufacturer's recommendation (Stem Cell Technology).

Overall organ assessments

For histology, various organs were fixed in 10% formalin, and 5 µM paraffin sections were stained with haematoxylin and eosin. Serum analysis for serum electrolytes, creatinine, calcium, AST, ALT, total protein, albumin and complete blood count (CBC) with differential were performed. Glucose tolerance tests (2 mg glucose per g body weight, intraperitoneal injection) were performed after 16 h fasting.

Immunohistochemistry

Ten-micrometre sections of the various formalin-fixed organs were prepared and standard immunohistochemical techniques were performed using a dilution of 1:500 tyrosine hydroxylase antibody (Chemicon) and CM5 p53 antibody (Nova-Castra). BrdU staining was performed accordingly on brain slices prepared from mice that had been continuously pulsed with BrdU at 4 h intervals for 24 h using a BrdU-staining kit purchased from Oncogene Research.

Neural stem-cell culture

Primary neural stem cells were isolated from the subventricular zone of age-matched adult mice using a method adapted from a previous study¹⁸. Deeply anaesthetized mice were cardiac-perfused with artificial cerebrospinal fluid (124 mM NaCl, 5 mM KCl, 1.3 mM MgCl₂, 2 mM CaCl₂, 26 mM NaHCO₃ and 10 mM D-glucose). Medial and lateral portions of the lateral ventricle subependyma were isolated from both brain hemispheres. Tissue was dissociated using trypsin (1.33 mg ml⁻¹; Sigma), hyaluronidase (Sigma) and kynurenic acid (Sigma). Trypsin was inactivated using serum-free trypsin inhibitor (Sigma). Cells were cultured in serum-free medium containing DMEM/F12 (Gibco), 5 mM HEPES buffer (Gibco), 0.6% glucose, 3 mM NaHCO₃, 2 mM glutamine, 25 µg ml⁻¹ insulin (Gibco), 100 mg ml⁻¹ transferrin (Gibco), 20 nM progesterone (Sigma), 60 µM putrescine (Gibco) and 30 mM sodium selenite (Gibco) in the presence of epidermal growth factor (20 mg ml⁻¹; Sigma). After 9–10 days in culture, we counted the number of neurospheres generated. Primary neurospheres were passaged by dissociation of the spheres into single cells using trituration through a fire-polished pipette.

Received 15 October; accepted 17 December 2002; doi:10.1038/nature01385.

Published online 22 January 2003.

- Shiloh, Y. & Kastan, M. B. ATM: genome stability, neuronal development, and cancer cross paths. *Adv. Cancer Res.* **83**, 209–254 (2001).
- Lee, H. W. *et al.* Essential role of mouse telomerase in highly proliferative organs. *Nature* **392**, 569–574 (1998).
- Blasco, M. A. *et al.* Telomere shortening and tumour formation by mouse cells lacking telomerase RNA. *Cell* **91**, 25–34 (1997).
- Borghesani, P. R. *et al.* Abnormal development of Purkinje cells and lymphocytes in *Atm* mutant mice. *Proc. Natl Acad. Sci. USA* **97**, 3336–3341 (2000).
- Kastan, M. B. & Lim, D. S. The many substrates and functions of ATM. *Nature Rev. Mol. Cell Biol.* **1**, 179–186 (2000).
- Metcalfe, J. A. *et al.* Accelerated telomere shortening in ataxia telangiectasia. *Nature Genet.* **13**, 350–353 (1996).
- Pandita, T. K. ATM function and telomere stability. *Oncogene* **21**, 611–618 (2002).
- Karseder, J., Broccoli, D., Dai, Y., Hardy, S. & de Lange, T. p53- and ATM-dependent apoptosis induced by telomeres lacking TRF2. *Science* **283**, 1321–1325 (1999).
- Barlow, C. *et al.* *Atm*-deficient mice: a paradigm of ataxia telangiectasia. *Cell* **86**, 159–171 (1996).
- Elson, A. *et al.* Pleiotropic defects in ataxia-telangiectasia protein-deficient mice. *Proc. Natl Acad. Sci. USA* **93**, 13084–13089 (1996).
- Xu, Y. *et al.* Targeted disruption of ATM leads to growth retardation, chromosomal fragmentation during meiosis, immune defects, and thymic lymphoma. *Genes Dev.* **10**, 2411–2422 (1996).
- Kipling, D. & Cooke, H. J. Hypervariable ultra-long telomeres in mice. *Nature* **347**, 400–402 (1990).
- Chin, L. *et al.* p53 deficiency rescues the adverse effects of telomere loss and cooperates with telomere dysfunction to accelerate carcinogenesis. *Cell* **97**, 527–538 (1999).
- Hande, M. P., Samper, E., Lansdorf, P. & Blasco, M. A. Telomere length dynamics and chromosomal instability in cells derived from telomerase null mice. *J. Cell Biol.* **144**, 589–601 (1999).
- Vaziri, H. *et al.* ATM-dependent telomere loss in aging human diploid fibroblasts and DNA damage lead to the post-translational activation of p53 protein involving poly(ADP-ribose) polymerase. *EMBO J.* **16**, 6018–6033 (1997).
- Hande, M. P., Balajee, A. S., Tchirkov, A., Wynshaw-Boris, A. & Lansdorf, P. M. Extra-chromosomal telomeric DNA in cells from *Atm*^{-/-} mice and patients with ataxia-telangiectasia. *Hum. Mol. Genet.* **10**, 519–528 (2001).
- Booth, C. & Potten, C. S. Gut instincts: thoughts on intestinal epithelial stem cells. *J. Clin. Invest.* **105**, 1493–1499 (2000).
- Reynolds, B. A. & Weiss, S. Clonal and population analyses demonstrate that an EGF-responsive mammalian embryonic CNS precursor is a stem cell. *Dev. Biol.* **175**, 1–13 (1996).
- Ahmed, S., Reynolds, B. A. & Weiss, S. BDNF enhances the differentiation but not the survival of CNS

- stem cell-derived neuronal precursors. *J. Neurosci.* **15**, 5765–5778 (1995).
20. Altman, J. & Bayer, S. A. in *Development of the Cerebellar System in Relations to its Evolution, Structure and Functions* 727–750 (CRC, London, 1997).
 21. Eilam, R. *et al.* Selective loss of dopaminergic nigro-striatal neurons in brains of Atm- deficient mice. *Proc. Natl Acad. Sci. USA* **95**, 12653–12656 (1998).
 22. Spector, B. D., Filipovich, A. H., Perry, G. S. & Kersey, J. H. in *Ataxia-telangiectasia—A Cellular and Molecular Link Between Cancer, Neuropathology and Immune Deficiency* (eds Bridges, B. A. & Harnden, D. G.) 103–138 (Wiley, New York, 1982).
 23. Rudolph, K. L. *et al.* Longevity, stress response, and cancer in aging telomerase-deficient mice. *Cell* **96**, 701–712 (1999).
 24. Artandi, S. E. *et al.* Telomere dysfunction promotes non-reciprocal translocations and epithelial cancers in mice. *Nature* **406**, 641–645 (2000).
 25. Tyner, S. D. *et al.* p53 mutant mice that display early ageing-associated phenotypes. *Nature* **415**, 45–53 (2002).
 26. Greenberg, R. A. *et al.* Short dysfunctional telomeres impair tumorigenesis in the INK4a(delta2/3) cancer-prone mouse. *Cell* **97**, 515–525 (1999).
 27. Smogorzewska, A. & de Lange, T. Different telomere damage signaling pathways in human and mouse cells. *EMBO J.* **21**, 4338–4348 (2002).
 28. Frank, K. M. *et al.* DNA ligase IV deficiency in mice leads to defective neurogenesis and embryonic lethality via the p53 pathway. *Mol. Cell* **5**, 993–1002 (2000).
 29. Sekiguchi, J. *et al.* Genetic interactions between ATM and the nonhomologous end-joining factors in genomic stability and development. *Proc. Natl Acad. Sci. USA* **98**, 3243–3248 (2001).
 30. Wong, K. K. *et al.* Telomere dysfunction impairs DNA repair and enhances sensitivity to ionizing radiation. *Nature Genet.* **26**, 85–88 (2000).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank D. Castrillon and M. Bosenberg for advice on histo-pathological analyses, and S. Weiler, D. Castrillon, N. Sharpless, N. Bardeesy, C. Khoo and S. Chang for critical reading of the manuscript. K.-K.W. is a Howard Hughes Physician Postdoctoral Fellow. R.S.M. is supported by a Damon Runyon Cancer Research Fund Fellowship. R.M.B. and D.R.C. are both NIH Mentored Clinician Scientists awardees. F.W.A. is supported by the Howard Hughes Medical Institute. This work was supported by grants from the NIH and ACS to R.A.D. R.A.D. is an American Cancer Society Research Professor and a Steven and Michele Kirsch Investigator.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.A.D.
(e-mail: ron_depinho@dfci.harvard.edu).

Plants lacking the main light-harvesting complex retain photosystem II macro-organization

A. V. Ruban^{*†}, M. Wentworth^{*†}, A. E. Yakushevska[‡], J. Andersson[§], P. J. Lee^{*}, W. Keegstra[‡], J. P. Dekker^{||}, E. J. Boekema[‡], S. Jansson[§] & P. Horton^{*}

^{*} Robert Hill Institute, Department of Molecular Biology and Biotechnology, University of Sheffield, Firth Court, Western Bank, Sheffield S10 2TN, UK

[‡] Biophysical Chemistry, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, Nijenborgh 4, 9747 AG Groningen, The Netherlands

[§] Umeå Plant Science Centre, Department of Plant Physiology, Umeå University, S-901 87 Umeå, Sweden

^{||} Faculty of Sciences, Division of Physics and Astronomy, Vrije Universiteit, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands

[†] These authors contributed equally to this work

Photosystem II (PSII) is a key component of photosynthesis, the process of converting sunlight into the chemical energy of life. In plant cells, it forms a unique oligomeric macrostructure in membranes of the chloroplasts¹. Several light-harvesting antenna complexes are organized precisely in the PSII macrostructure—the major trimeric complexes (LHCII)² that bind 70% of PSII chlorophyll and three minor monomeric complexes³—which together form PSII supercomplexes^{4–6}. The antenna complexes are essential for collecting sunlight and regulating photosyn-

thesis^{7–9}, but the relationship between these functions and their molecular architecture is unresolved. Here we report that anti-sense *Arabidopsis* plants lacking the proteins that form LHCII trimers¹⁰ have PSII supercomplexes with almost identical abundance and structure to those found in wild-type plants. The place of LHCII is taken by a normally minor and monomeric complex, CP26, which is synthesized in large amounts and organized into trimers. Trimerization is clearly not a specific attribute of LHCII. Our results highlight the importance of the PSII macrostructure: in the absence of one of its main components, another protein is recruited to allow it to assemble and function.

The central unit of PSII is the core complex, in which light is used to oxidize water to molecular oxygen, to reduce plastoquinone and to generate a transmembrane proton gradient^{11,12}. In plants, each core dimer binds two copies of the minor, monomeric light-harvesting proteins CP29, CP26 and CP24 (the *Lhcb4*, *Lhcb5* and *Lhcb6* gene products, respectively) and two–four LHCII trimers (made up from the *Lhcb1*, *Lhcb2* and *Lhcb3* gene products) to form a PSII supercomplex^{13,14}. Additional LHCII trimers have been located in membrane domains deficient in PSII supercomplexes, although they still can be involved in efficient light-harvesting¹³. These trimers consist of the *Lhcb1* and *Lhcb2* gene products³, and can partition, depending on the phosphorylation state of LHCII, between the grana and stromal thylakoids to transfer energy to PSII and PSI, respectively^{15,16}.

This model for PSII macrostructure suggests that each light-harvesting protein has a unique position and role, but little is known about the structure–function relationships involved. In addition, the possibility cannot be excluded that there is some redundancy

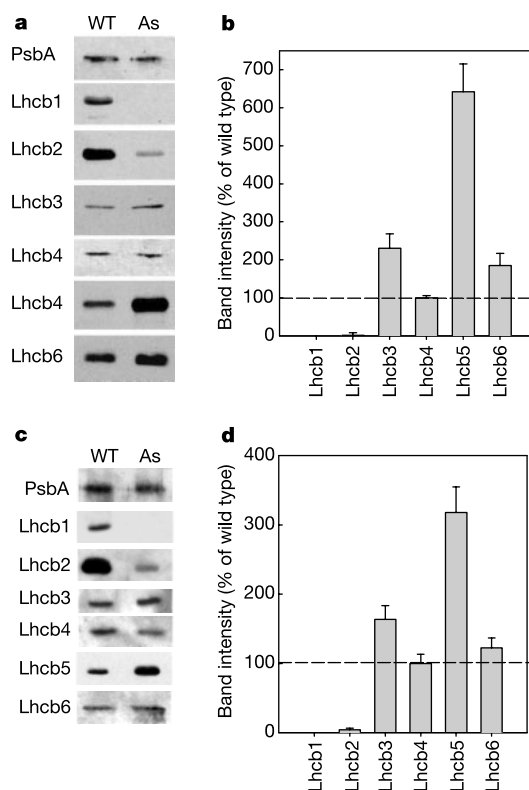


Figure 1 PSII membrane composition. **a, c**, Western blots of PSII membranes (**a**) and supercomplexes (**c**) prepared from wild-type (WT) and asLhcb2 (As) plants. Gels were probed with antibodies specific for the proteins indicated on the left. Sample loading was done on an equal chlorophyll basis. **b, d**, Results of densitometric analysis of the western blots of PSII membranes (**b**) and supercomplexes (**d**). Values are normalized to the amount of PsbA and are the means $\pm$ s.e.m. of four replicate gels.

among these similar light-harvesting proteins¹⁷. We have taken a genetic approach to investigating these complexes. Expression of a full-length *Lhcb2* antisense construct in *Arabidopsis thaliana* abolishes the expression of both *Lhcb2* and *Lhcb1*, owing to stretches of identical nucleotide sequences in these two genes¹⁰. The plants (called asLhcb2) are almost completely devoid of the two main proteins of trimeric LHCII, Lhcb1 and Lhcb2 (ref. 10). Western blot analysis of PSII membranes prepared from wild-type and asLhcb2 plants showed that Lhcb1 is completely absent and the amount of Lhcb2 is less than 5% of that found in wild type (Fig. 1a). When normalized to the amount of the PSII core protein PsbA, the quantities of Lhcb3, Lhcb4 and Lhcb6 were constant or less than twofold higher, whereas the amount of Lhcb5, the apoprotein of CP26, increased by more than sixfold (Fig. 1b).

The absorption spectra showed a strong reduction in the chlorophyll *b* band at 650 nm in the antisense membranes (Fig. 2a, curves 1 and 2), which is consistent with the measured chlorophyll *a/b* ratio of 4.3, as compared with 3.3 in the wild type, and is explained by the depletion of LHCII and the increase in CP26 (CP26 has a considerably higher chlorophyll *a/b* ratio than has LHCII¹⁸). The similarity between the fluorescence excitation spectrum and the absorption spectrum in both membranes (Fig. 2a, dotted lines) shows that the extra CP26 functions as an efficient light-harvesting antenna for PSII. Together with the identical photosynthetic activity of the wild-type and antisense plants¹⁰, this indicates that PSII activity is the same in the wild-type and antisense plants.

Chromatographic analysis of detergent-solubilized PSII membranes¹⁹ gave unexpected results. Fraction II, containing PSII super-

complexes, was still present in the antisense plants (Fig. 2b) but was highly enriched in CP26, containing over three times more Lhcb5 than present in the wild-type fraction (Fig. 1c, d). The quantity of Lhcb3 was also higher, whereas that of Lhcb4 and Lhcb6 was unchanged. These data indicate that the missing Lhcb1 and Lhcb2 in the supercomplex are replaced by CP26, which prompts an

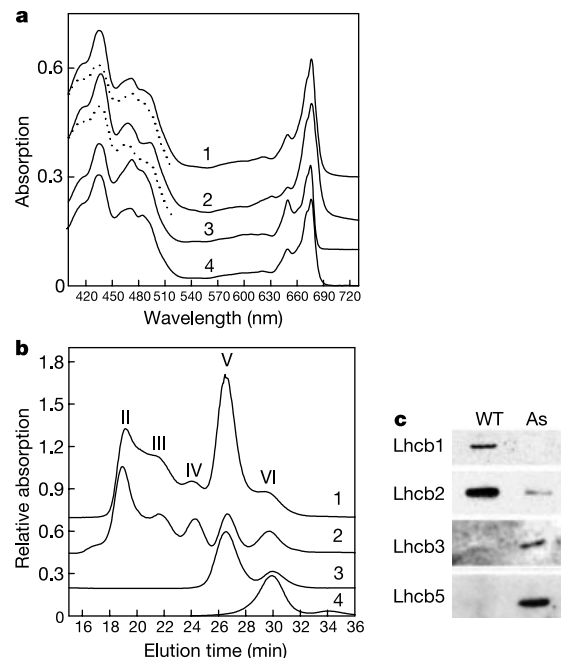


Figure 2 PSII membrane spectra. **a**, Absorption spectra of PSII membranes (curves 1 and 2) and gel-filtration fraction V (curves 3 and 4) prepared from wild-type (curves 1 and 3) and asLhcb2 (curves 2 and 4) plants. Dotted spectra show a fragment of the excitation spectrum of PSII fluorescence emission (685 nm) for the PSII membrane samples. **b**, Gel-filtration elution profiles of solubilized PSII membranes from wild-type (1) and asLhcb2 (2) samples, detected at 670 nm. The main fractions contain PSII supercomplexes (II), PSI (III), PSII cores (IV), trimeric light-harvesting complexes (V) and monomeric light-harvesting complexes (VI). Curves 3 and 4 show the elution profiles of purified LHCII trimers and a monomeric LHC fraction. **c**, Western blots of fraction V prepared from wild-type (WT) and asLhcb2 (As) plants. Details are as described in Fig. 1.

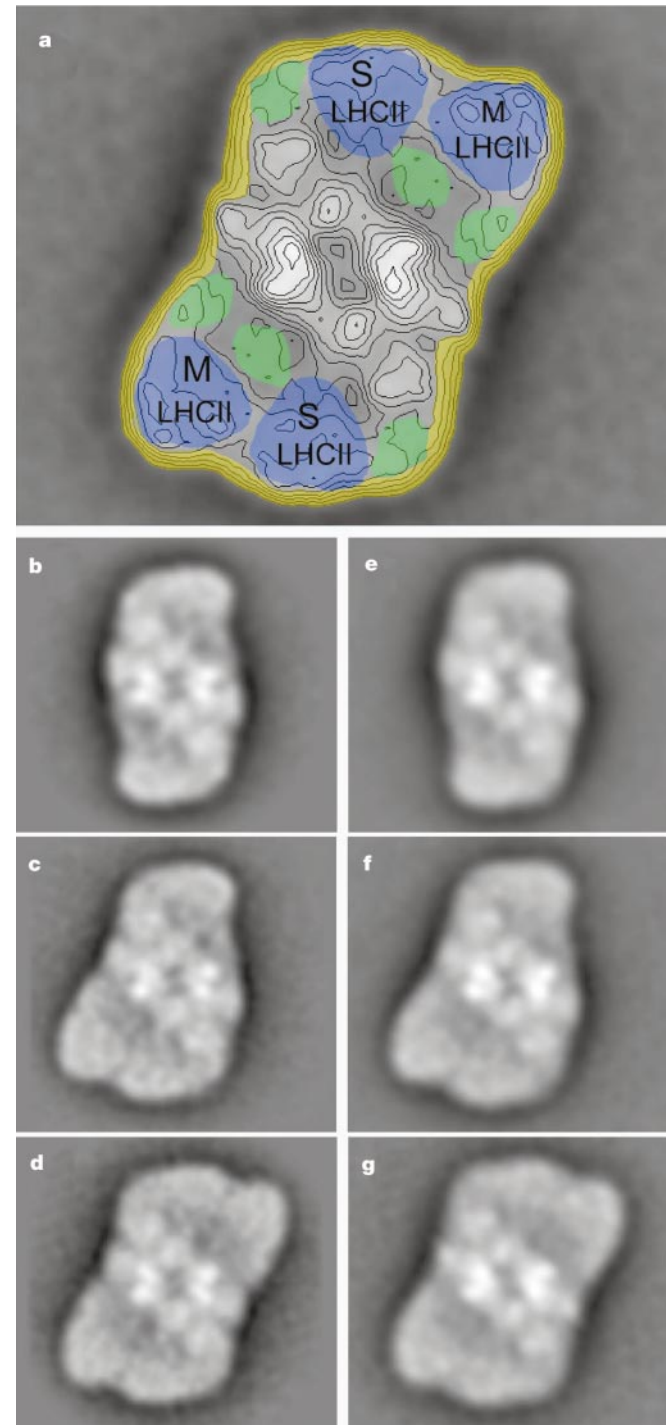


Figure 3 Average projections of the PSII complexes from *Arabidopsis* Lhcb2 plants. **a**, An *Arabidopsis* C₂S₂M₂ supercomplex¹⁹ obtained previously, depicting the S- and M-LHCII trimers (blue), the monomeric complexes (green) and the detergent shell (yellow). **b–d**, Antisense Lhcb2 plants; **e–g**, wild-type plants. **b**, Average of 510 C₂S₂ projections; **c**, average of 290 C₂S₂M projections; **d**, average of 411 C₂S₂M₂ projections; **e**, average of 300 C₂S₂ projections; **f**, average of 948 C₂S₂M projections; **g**, average of 218 C₂S₂M₂ projections.

intriguing question: can CP26 assemble itself into a trimeric form to mimic the LHCII trimer? In fact, the gel-filtration separation procedure showed that a trimeric fraction (Fig. 2b, fraction V) was still present in the antisense plants, although its concentration was strongly reduced. Western blot analysis of this trimer showed that it contained only CP26 and Lhcb3 (Fig. 2c), whereas the wild-type fraction V contained almost exclusively Lhcb2 and Lhcb1, with some Lhcb3 (but at a much lower amount than in the antisense plant), as expected (refs 3, 7, and Fig. 2c).

The absorption spectra of the wild-type and antisense trimers were also clearly different (Fig. 2a (curves 3 and 4)), with reduced absorption at 650 nm from chlorophyll *b* in the antisense trimer. We calculated the chlorophyll *a/b* ratio of the wild-type trimer to be 1.4 as expected²³, but this ratio increased to 1.90–1.95 in the antisense plant. This higher ratio is explained by the chlorophyll *a/b* ratios of CP26 (2.0)¹⁸, which we confirmed to be the same in the CP26 from both wild-type and antisense plants, and Lhcb3 (1.75)²⁰. The ability to assemble into trimers *in vivo* was previously thought to be specific for Lhcb1 and Lhcb2. Clearly, this is not so.

Of all the other Lhc genes, *Lhcb5* bears the strongest sequence similarity to *Lhcb1* and *Lhcb2* (ref. 17), greater than 40%, as compared with roughly 30% or less for the other Lhc genes. *Lhcb5* is the only gene that clusters with the *Lhcb1*–*Lhcb3* genes¹⁷. In addition, a WYXXR motif near the amino terminus of Lhcb1, which is necessary for LHCII trimerization²¹, is conserved in Lhcb5, the apoprotein of CP26, but not in the apoproteins of CP29 and CP24. But another LHCII trimerization domain near the carboxy terminus (Trp 222)²² is not found in Lhcb5, which could explain the reduced efficiency of trimerization for CP26.

We also analysed the isolated supercomplex population by electron microscopy and image analysis. The same three main classes of supercomplex⁴ were found in both asLhcb2 (Fig. 3b–d) and wild-type (Fig. 3e–g) samples. Clearly, there are trimers lacking Lhcb1 and Lhcb2 that are assembled into the supercomplex in the antisense plants and, within the 2-nm resolution of the negative staining, are identical to those found in the wild type. Notably, the abundance of supercomplexes was about the same in both samples. Thus, they cannot arise from the residual Lhcb2 present in the antisense plants and cannot be due to Lhcb3, because the amount of this protein is not increased.

To find out how the supercomplexes are organized in the thylakoid membrane, electron microscopy was carried out on grana membrane fragments^{13,19}. Ordered arrays of PSII supercomplexes were found in both wild-type and antisense plants (Fig. 4a, b).

The frequency of occurrence of these arrays was judged to be about the same and, notably, the structural features were almost identical. Both images show arrays of PSII supercomplexes (Fig. 4b, red outline)¹⁹. In each one, rows of PSII cores (bright areas) are separated by areas of trimers, despite the absence in the antisense plants of the main proteins that are known to form LHCII trimers in the wild-type plants. The spacing between the rows is slightly reduced in the antisense plants, suggesting a slightly different packing of the complexes. Notably, also in the membranes from the antisense plants, the contours of the trimers are clearly apparent (Fig. 4a, arrows). Again, within the resolution of the technique, these trimers appear to be identical to those seen in the wild type.

These data show that the PSII supercomplexes have a precise molecular design in which trimers enable a normal supercomplex to be assembled and allow its correct organization in the thylakoid membrane. The observations that ordered arrays of PSII are formed in the antisense plants and that other proteins have been recruited to allow this formation show the importance of this feature of macromolecular organization of PSII. The plants attempt to compensate for the absence of LHCII by synthesizing more CP26 and assembling it into trimers, which then function as a PSII antenna and enable a normal macro-organization to be attained, including their grana structure¹⁰. This organization is essential for the high efficiency of energy usage in photosynthesis, because it allows energy exchange between PSII cores and energy migration over a large domain of chlorophyll molecules²³.

Because the wild-type supercomplex contains two CP26 copies for each dimeric PSII core complex, we estimate from the western blot analysis that the supercomplex from the antisense plants contains at least four additional copies of CP26 and up to two additional copies of Lhcb3. Thus, there is sufficient CP26 and Lhcb3 to form at least two trimers, thereby explaining how the structure of the complex can be virtually identical to the wild-type even though Lhcb1 and Lhcb2 are absent. In the PSII membranes, in which there is a sixfold increase in CP26, we estimate the presence of two further CP26-containing trimers, explaining the presence of C₂S₂M₂ supercomplexes.

The compensatory replacement of LHCII by CP26 does not give rise to a PSII unit of identical structure and function. There are small differences in the packing of the supercomplexes in the grana membrane but, more notably, the extra LHCII trimers per PSII not present in the supercomplex in wild-type plants¹³ are apparently not replaced by CP26 in the antisense plants. Thus, there is a reduction in the total content of light-harvesting pigments. In addition, CP26 does not have the capacity to be phosphorylated,

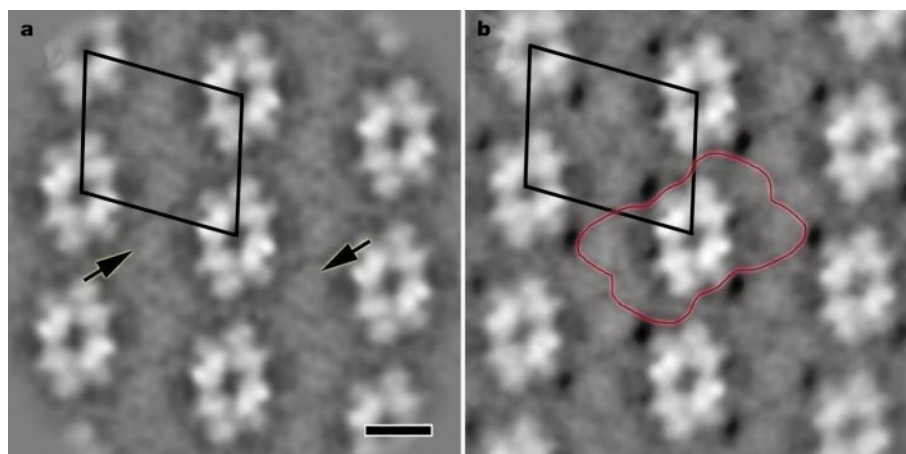


Figure 4 Final result of image analysis of two-dimensional crystalline PSII complexes from *Arabidopsis* wild-type and Lhcb2 antisense plants. **a**, Sum of 600 aligned crystal fragments from *Arabidopsis* Lhcb2 antisense plant with the unit cell (24.0 × 21.8 nm)

and two trimers (arrows) indicated. **b**, Sum of 450 aligned crystal fragments from *Arabidopsis* wild-type with the unit cell (25.6 × 21.4 nm) and an outline of the fitting of the C₂S₂M₂ supercomplex (red line) indicated. Scale bar, 10 nm.

and therefore the antisense plants cannot carry out state transitions¹⁰, a regulatory process that balances the excitation of the two photosystems and requires LHCII phosphorylation^{15,16}. These two differences almost certainly account for the phenotype of these plants, the reduced fitness in the field¹⁰ and the failure to grow normally under conditions of very low irradiance (J. Andersson and S. Jansson, unpublished data). Although there is some apparent redundancy between the similar Lhcb proteins, each one has specific features that confer its unique role in PSII structure and function. Lhcb1 and Lhcb2 are designed to form the essential trimeric building blocks of the supercomplex, where they form the major part of the PSII antenna, but they also provide the vital flexibility needed to regulate excitation distribution between the photosystems. When Lhcb1 and Lhcb2 are missing, CP26 can substitute but not completely. This explains why normally the precision of the composition of the supercomplexes is maintained absolutely in spite of this potential redundancy.

Do other membrane protein complexes show similar features? In cyanobacterial photosynthetic membranes, the CP43' protein, which closely resembles the PSII core protein CP43, is synthesized under certain stress conditions and forms an oligomeric ring-like structure around PSI^{24,25}. Thus, there are hitherto unrecognized features of the structure of macromolecular membrane protein complexes: the same or similar proteins may adopt different structural and functional roles in different or even the same complexes; and complexes that have the same overall structure may contain different proteins resulting in different functions. □

Methods

Plant growth

Arabidopsis thaliana cv. Columbia plants were grown in growth chambers with an 8-h photoperiod and 200 $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$ for 8 weeks as described²⁶. The *asLhcb2* line used shows almost total absence of the *Lhcb1* and *Lhcb2* gene products¹⁰.

Sample preparation and analysis

PSII membranes were prepared as previously described²⁷, but with some modifications. In brief, leaves were homogenized three times for 5 s in a medium containing 20 mM Tricine-NaOH, pH 8.4, 0.45 M sorbitol, 10 mM EDTA and 0.1% bovine serum albumin. After pelleting and washing in 0.3 M sorbitol, 20 mM Tricine NaOH, pH 7.6, 5 mM MgCl₂, chloroplasts were osmotically shocked for 30 s with 5 mM MgCl₂, pH 7.6. Pelleted thylakoids were incubated in stacking medium (5 mM MgCl₂, 15 mM NaCl and 2 mM MES, pH 6.3) for 1 h, after which we treated them with 2.5% Triton X-100 at a chlorophyll concentration of 3 mg ml⁻¹. The PSII membranes were sedimented at 30,000g for 30 min, resuspended in buffer containing 20 mM Bis-Tris (pH 6.5) and 5 mM MgCl₂, and solubilized with 0.4% *n*-dodecyl- α -D-maltoside at a chlorophyll concentration of 1.4 mg ml⁻¹. Large fragments were removed by centrifugation for 3 min at 9,000 r.p.m. and the supernatant was filtered promptly through a 0.45- μm filter. Finally, the sample was subjected to gel-filtration chromatography using a Superdex 200 HR 10/30 column in an Amersham-Pharmacia Äcta Purifier system using the same buffer and similar conditions as before¹⁹. We prepared purified trimeric LHCII and monomeric LHC from spinach PSII membranes²⁸.

Spectroscopic analysis

Low-temperature spectroscopy was done using an Optistat^{DN} LN-2 cooled bath cryostat (Oxford Instruments). Samples were diluted in a medium containing 70% glycerol (w/v), 20 mM HEPES buffer, pH 7.8, 5 mM MgCl₂ and 0.33 M sorbitol. The chlorophyll concentration for absorption was 4 μM and for the fluorescence excitation measurements was 1 μM . A poly(methyl metacrylate) PMMA 1-cm cuvette was used for all experiments. Absorption measurements were carried out using a Cary 500 UV-Vis-NIR spectrophotometer and corrected fluorescence excitation spectra were recorded using a SPEX FluoroLog FL3-22 spectrofluorimeter.

Electrophoresis and western blot analysis

Protein samples were solubilized and separated by 15% denaturing SDS-PAGE essentially as described²⁹. Roughly equal amounts of chlorophyll were loaded, about 2 μg per lane for PSII membranes and 0.5 μg for isolated supercomplexes and trimers. Proteins were transferred to Hybond-P PVDF membrane (Amersham Pharmacia) in a Mini-Trans-Blot transfer cell (Bio-Rad) at 30 mA for 12 h. Membranes were detected with specific antibodies against the proteins Lhcb1–Lhcb6 (ref. 30) and PsbA (a gift from P. Nixon). The primary antibody was detected by a horseradish peroxidase (HRP)-labelled secondary antibody using an ECL Plus kit (Amersham Pharmacia). Chemiluminescence was detected on Hyperfilm ECL (Amersham Pharmacia) photographic film. We developed the films for 20 min, dried and digitized them in 256-bit greyscale at a resolution of 600 d.p.i. using an Umax Powerlook III high-resolution scanner set in transmission mode. (Films were developed for 0.5–60 min and densities were found to be linear over the

first 40 min for each antibody.) A standard Kodak photographic 21 step tablet (OD 0.05–3.05) was used to calibrate the scanner each time an image was scanned. We processed images using the 1D software package of the Image Master Suite (Amersham Pharmacia). Individual lanes and bands were detected automatically by the software, and background for each lane was subtracted using the rolling disc option to give the final band density. For the antibodies against Lhcb1, Lhcb2 and Lhcb5, linear densitometric responses were found in the range of 0.5–10 μg of chlorophyll per lane for PSII membranes.

Electron microscopy

After chromatography, fractions were immediately prepared for electron microscopy. Samples were negatively stained using the droplet method with 2% uranyl acetate and were prepared on glow discharge carbon-coated copper grids as described⁴. The membrane fragments and supercomplexes were analysed by a similar method to that used for preparations from spinach and the wild-type *Arabidopsis* plants^{13,19}. The samples were imaged in a Philips CM10 electron microscope at $\times 52,000$ magnification. Electron micrographs were digitized with a Kodak Eikonix Model 1412 CCD (charge-coupled device) camera. Single-particle projections were extracted from negatives and analysed with IMAGIC software (Image Science Software GmbH) and Groningen Image Processing ('GRIP') software (Groningen University Software)^{13,19}.

Received 22 August; accepted 3 December 2002; doi:10.1038/nature01344.

- Hankamer, B., Barber, J. & Boekema, E. J. Structure and membrane organisation of photosystem II in green plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**, 641–671 (1997).
- Kühlbrandt, W. & Wang, D. N. Three dimensional structure of plant light-harvesting complex by electron crystallography. *Nature* **350**, 130–134 (1991).
- Peter, G. F. & Thornber, J. P. Biochemical composition and organization of higher plant photosystem II light-harvesting pigment-proteins. *J. Biol. Chem.* **266**, 16745–16754 (1991).
- Boekema, E. J., Van Roon, H., Calkoen, F., Bassi, R. & Dekker, J. P. Multiple types of association of photosystem II and its light-harvesting antenna in partially solubilized photosystem II membranes. *Biochemistry* **38**, 2233 (1999).
- Boekema, E. J., Van Roon, H., Van Breemen, J. F. L. & Dekker, J. P. Supramolecular organization of photosystem II and its light-harvesting antenna in partially solubilized photosystem II membranes. *Eur. J. Biochem.* **266**, 444–452 (1999).
- Nield, J., Orlova, E. V., Morris, E. P., Gowen, B., Van Heel, M. & Barber, J. 3D map of the plant photosystem II supercomplex obtained by cryoelectron microscopy and single particle analysis. *Nature Struct. Biol.* **7**, 44–47 (2000).
- Jansson, S. The light-harvesting chlorophyll *a/b*-binding proteins. *Biochim. Biophys. Acta* **1184**, 1–19 (1994).
- Paulsen, H. Chlorophyll *a/b* binding proteins. *Photochem. Photobiol.* **62**, 367–382 (1995).
- Horton, P., Ruban, A. V. & Walters, R. G. Regulation of light harvesting in green plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **47**, 655–684 (1996).
- Andersson, J. *et al.* Absence of the main light harvesting complex of photosystem II affects photosynthetic function. *Plant J.* (in the press).
- Barber, J. Photosystem II: a multisubunit membrane protein that oxidizes water. *Curr. Opin. Struct. Biol.* **12**, 523–530 (2002).
- Zouni, A. Crystal structure of photosystem II from *Synechococcus elongatus* at 3.8 Å resolution. *Nature* **409**, 739–743 (2001).
- Boekema, E. J., Van Breemen, J. F. L., Van Roon, H. & Dekker, J. P. Arrangement of photosystem II in crystalline macrodomains within the thylakoid membrane of green plant chloroplasts. *J. Mol. Biol.* **301**, 1123–1133 (2000).
- Hankamer, B. *et al.* Isolation and characterization of monomeric and dimeric photosystem II complexes from spinach and their relevance to the organisation of photosystem II *in vivo*. *Eur. J. Biochem.* **243**, 422–429 (1997).
- Allen, J. F. & Forsberg, J. Molecular recognition in thylakoid structure and function. *Trends Plant Sci.* **6**, 317–326 (2001).
- Haldrup, A., Jensen, P. E., Lunde, C. & Scheller, H. V. Balance of power: a view of the mechanism of photosynthetic state transitions. *Trends Plant Sci.* **6**, 301–305 (2001).
- Jansson, S. A guide to the *Lhc* genes and their relatives in *Arabidopsis*. *Trends Plant Sci.* **4**, 236–240 (1999).
- Croce, R., Canino, G., Ros, F. & Bassi, R. Chromophore organisation in the higher plant photosystem II antenna protein CP26. *Biochemistry* **41**, 7334–7343 (2002).
- Yakushevskaya, A. E. *et al.* Supramolecular organisation of photosystem II and its associated light harvesting antenna in *Arabidopsis thaliana*. *Eur. J. Biochem.* **268**, 6020–6028 (2001).
- Caffrari, S., Croce, R., Cattivelli, L. & Bassi, R. The Lhcb1, 2 and 3 gene products, components of the trimeric antenna complex of higher plant photosystem II, have distinct biochemical and spectroscopic properties. *Proc. 12th Int. Congr. Photosyn.*, 531–034 (CSIRO, Canberra, 2001).
- Hobe, S., Foster, R., Klingler, J. & Paulsen, H. N-proximal sequence motif in light-harvesting chlorophyll-*a/b*-binding protein is essential for trimerisation of the light harvesting chlorophyll *a/b* complex. *Biochemistry* **34**, 10224–10228 (1995).
- Kuttkat, A., Kartmann, A., Hobe, S. & Paulsen, H. The C-terminal domain of light-harvesting chlorophyll-*a/b*-binding protein is involved in the stabilisation of trimeric light harvesting complex. *Eur. J. Biochem.* **242**, 288–292 (1996).
- Garab, G. & Mustardy, L. Role of LHCII-containing macrodomains in the structure, function and dynamics of grana. *Aust. J. Plant Physiol.* **27**, 648–658 (1999).
- Bibby, T., Nield, J. & Barber, J. Iron deficiency induces the formation of an antenna ring around trimeric photosystem I in cyanobacteria. *Nature* **412**, 743–745 (2001).
- Boekema, E. J. *et al.* A giant chlorophyll-protein complex induced by iron deficiency in cyanobacteria. *Nature* **412**, 745–748 (2001).
- Walters, R. G., Rogers, J. J. M., Shephard, F. & Horton, P. Acclimation of *Arabidopsis thaliana* to the light environment: the role of photoreceptors. *Planta* **209**, 517–527 (1999).
- Berthold, D. A., Babcock, G. T. & Yocum, C. F. A highly resolved, oxygen-evolving photosystem II preparation from spinach thylakoid membranes. EPR and electron transport properties. *FEBS Lett.* **134**, 231–234 (1981).

28. Ruban, A. V., Lee, P. J., Wentworth, M., Young, A. J. & Horton, P. Determination of the stoichiometry and strength of binding of xanthophylls to the photosystem II light harvesting complexes. *J. Biol. Chem.* **274**, 10458–10465 (1999).
29. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).
30. Andersson, J., Walters, R. G., Horton, P. & Jansson, S. Antisense inhibition of the photosynthetic antenna proteins CP29 and CP26: implications for the mechanism of protective energy dissipation. *Plant Cell* **13**, 1193–1204 (2001).

Acknowledgements We wish to thank R. Walters for discussions. This work was supported by the UK Biotechnology and Biological Sciences Research Council, the UK Joint Infrastructure Fund, the Netherlands Foundation for Scientific Research (NWO) through the Foundation for Life and Earth Sciences (ALW), and the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning and the Foundation for Strategic Research.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.H. (e-mail: p.horton@sheffield.ac.uk).

Structure and catalytic mechanism of the human histone methyltransferase SET7/9

Bing Xiao*, Chun Jing*, Jonathan R. Wilson*, Philip A. Walker*, Nishi Vasisht*, Geoff Kelly*, Steven Howell*, Ian A. Taylor*, G. Michael Blackburn† & Steven J. Gamblin*

* Structural Biology Group, National Institute for Medical Research, Mill Hill, London NW7 1AA, UK

† Department of Chemistry, Krebs Institute, University of Sheffield, Sheffield S3 7HF, UK

Acetylation^{1,2}, phosphorylation³ and methylation⁴ of the amino-terminal tails of histones are thought to be involved in the regulation of chromatin structure and function^{5–7}. With just one exception^{8,9}, the enzymes identified in the methylation of specific lysine residues on histones (histone methyltransferases) belong to the SET family¹⁰. The high-resolution crystal structure of a ternary complex of human SET7/9 with a histone peptide and cofactor reveals that the peptide substrate and cofactor bind on opposite surfaces of the enzyme. The target lysine accesses the active site of the enzyme and the *S*-adenosyl-*L*-methionine (AdoMet) cofactor by inserting its side chain into a narrow channel that runs through the enzyme, connecting the two surfaces. Here we show from the structure and from solution studies that SET7/9, unlike most other SET proteins, is exclusively a mono-methylase. The structure indicates the molecular basis of the specificity of the enzyme for the histone target, and allows us to propose a model for the methylation reaction that accounts for the role of many of the residues that are invariant across the SET family.

Many SET proteins have now been characterized biochemically and several have been the subject of X-ray structure analysis: SET7/9 from human¹¹ and its complex with the product *S*-adenosyl-*L*-homocysteine (AdoHcy)¹²; Dim-5 from *Neurospora crassa*⁶; Rubisco large subunit methyltransferase (LSMT) from pea with AdoHcy¹³; and Ctr4 from *Schizosaccharomyces pombe*¹⁴. SET proteins can be classified according to the lysine residues that they target on histones H3, H4 and H2A⁴, and it is apparent that methylation at these different sites gives rise to distinct biological effects. An additional level of complexity is that lysine residues may be mono-, di- or tri-methylated and that these distinct species lead

to different signalling events. For example, in *Saccharomyces cerevisiae*, although di-methylation of Lys 4 on histone H3 is present at both active and inactive euchromatic genes, tri-methylation is linked exclusively to active genes¹⁵.

NMR studies (see Supplementary Information) indicated that a histone peptide containing mono-methylated Lys 4 was better ordered in complex with SET7/9 than unmodified peptide. We therefore used the products of the normal histone methyltransferase (HMT) reaction for crystallization experiments (methylated lysine peptide and AdoHcy). In our previous studies¹¹ we obtained useful SET7/9 crystals only from constructs lacking the small carboxy-terminal segment, which is, nevertheless, essential for the catalytic activity of the enzyme. Here, we have used a catalytically active construct that contains the complete C-terminal segment. We obtained well-ordered crystals of the ternary complex of SET7/9 that diffracted to at least 1.7 Å spacing, and the structure was readily solved by molecular replacement. The C-terminal segment, the AdoHcy cofactor and most of the substrate peptide are well defined in the electron density maps, as are all the important residues around the active site. The overall structure of the ternary complex

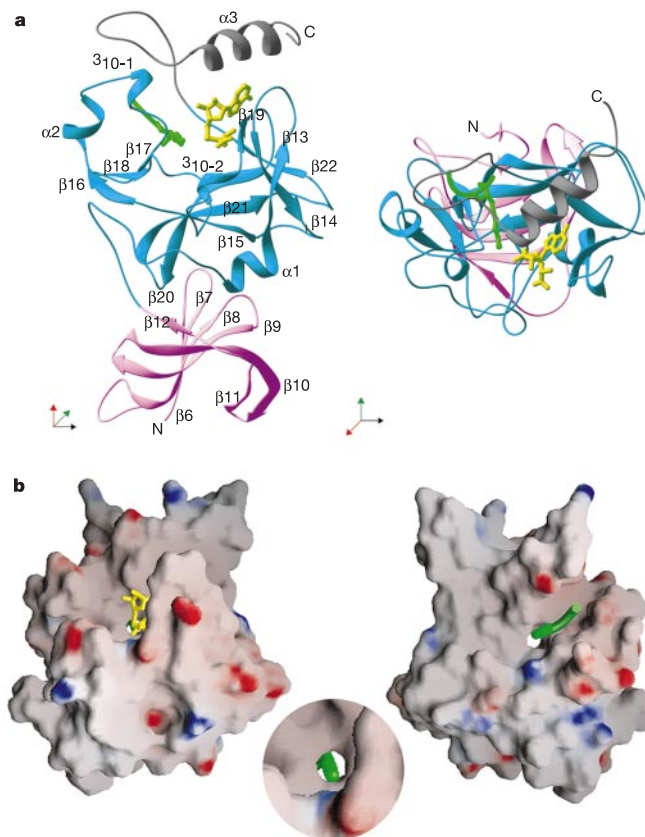


Figure 1 Structure of the SET7/9 ternary complex. **a**, Two orthogonal views of the SET7/9 ternary complex in ribbons representation. The N-terminal domain is coloured pink, the SET domain is blue and the C-terminal segment is grey. The H3 peptide is indicated in green, with the side chain of methylated Lys 4 shown. The *S*-adenosyl-*L*-homocysteine (AdoHcy) cofactor is coloured yellow. The secondary structure elements are labelled according to our earlier structure. Two small turns of the 3₁₀ helix are also labelled. **b**, Two views of the SET domain are shown in a surface representation coloured according to electrostatic potential (the two views are related by a twofold rotation about a vertical axis). The left panel shows AdoHcy coloured yellow; the right panel shows the H3 peptide coloured green. The inset panel shows a close-up view of the lysine access channel containing the methyl lysine side chain as viewed from the *S*-adenosyl-*L*-methionine (AdoMet)-binding site.

of SET7/9 is shown in Fig. 1 (see Supplementary Information for crystallographic statistics).

The most notable feature of the catalytic structure is that the AdoHcy and the peptide substrate are located on opposite sides of the SET domain and that there is a narrow channel passing through the enzyme that connects the peptide and cofactor binding surfaces. The target lysine residue of the substrate (Lys 4) is inserted into this channel so that its amine can access the methyl donor (AdoMet; see Fig. 1b). The packing of the C-terminal segment against the SET domain is required to form the lysine access channel, which explains

why this feature has not been observed in previous HMTase structures (although it was noted in the context of Rubisco LSMT, as was the presence of a molecule of HEPES in the channel)¹³. The C-terminal segment (residues 345–366) is organized into two structural features. Residues 337–349, belonging mainly to the SET domain, form an approximate β -hairpin structure that protrudes at a right angle to the surface of the enzyme. This is followed by three residues that accommodate a sharp bend in the polypeptide chain before the final stretch of the protein that adopts an α -helical conformation (Fig. 1a). Tyr 335 and Tyr 337, located

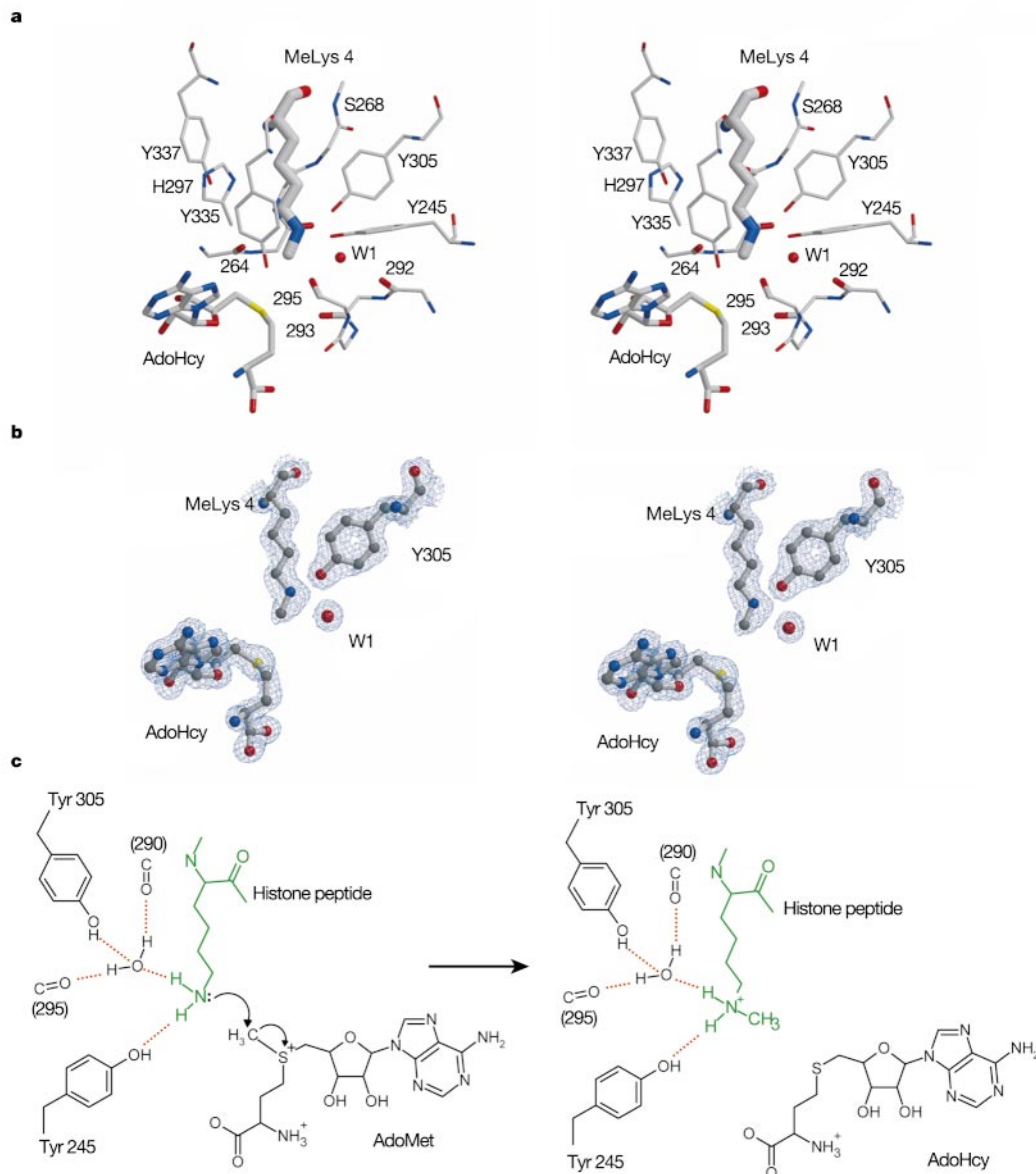


Figure 2 Active site of SET7/9. **a**, Stereographic representation of selected active-site residues together with methylated Lys 4 (MeLys 4), AdoHcy and several main-chain carbonyl groups oriented towards the lysine amine. **b**, Electron density ($2F_o - F_c$) covering part of the active site; the view is in the same orientation as in **a**. **c**, Schematic diagram of the reaction. The structure shows that the lysine has been stripped of all solvent molecules except for the one water used as a hydrogen-bond acceptor to orient the amino group. This desolvation will lower the pK_a of the lysine amino group and also enhance its nucleophilicity. At the same time, the local orientation of the dipoles of main-chain carbonyl groups towards the nitrogen will stabilize the developing positive charge on

that atom as the methylation reaction proceeds. There is no proximate proton acceptor to provide general base catalysis for the nucleophilic nitrogen. Mechanistically, it seems probable that the lysine side chain enters the active site with difficulty in its protonated form, the passage of this cation through the channel being facilitated by the faces of the flanking tyrosines. In the active site, the desolvated lysine is deprotonated, possibly to one of the flanking tyrosine oxygens. Thereafter the methylation reaction proceeds without general base catalysis, facilitated simply by alignment of orbitals, by desolvation, and by stabilization of charge reorganization.

just before the C segment, are both important for the formation of the lysine access channel. The arrangement of the β -hairpin is such that it stabilizes the conformation of these two tyrosine residues while also contributing to one of the sides of the groove into which the peptide binds. The second side of the peptide-binding groove is made up by residues 255–268 (including β -17). The α -helix at the end of the C-terminal segment packs against β -19, particularly Phe 299 (located just beyond the conserved NHS signature motif)¹⁶, and makes hydrophobic packing interactions with the AdoHcy cofactor through Trp 352.

From the high-resolution structure presented here, and from the previous co-crystal structures of Rubisco LSMT/AdoHcy¹³ and

SET7/AdoHcy¹², it is apparent that in our earlier studies the orientation of AdoMet, which we modelled into a low-resolution difference map, was incorrect. The mode of cofactor binding in our present structure is essentially identical to that described previously^{12,13}. In this alignment, the methyl group to be transferred from the AdoMet to the amine is pointing into the lysine-binding channel (see Fig. 2a). At the peptide-binding site, the channel surface is largely made up by the side chains of Leu 267 and Tyr residues 305, 335 and 337 (Fig. 2a). The alkyl component of the lysine side-chain therefore inhabits a hydrophobic environment. The other end of the channel, where it opens onto the cofactor-binding surface, is dominated by four tyrosine OH groups (Tyr 245, 305, 335, 337) and five main-chain carbonyl groups, all approximately oriented towards the lysine amine group. There are also several well-defined water molecules in the vicinity of the active site, although only one is close to the lysine amine. The structure, with its catalytic and binding surfaces, explains the role of a number of invariant residues whose mutation to alanine has been shown to abolish HMTase activity without significantly affecting the affinity for substrate or cofactor binding. The aromatic ring of Tyr 335 forms a significant part of the binding cavity for the alkyl portion of the Lys 4_{PEP} residue (histone peptide substrate) while its side-chain oxygen makes a single hydrogen bond to the main-chain carbonyl of residue 295. The importance of the correct positioning of this tyrosine residue is underscored by the observation that His 297 (of the NHS signature) acts to stabilize its conformation by acting as a hydrogen-bond acceptor to its main-chain amide (Fig. 2a).

The electron density for the methyl Lys 4_{PEP} is very well defined and clearly shows the location of the single methyl group (Fig. 2b). Examination of the active site also reveals that the lysine amine

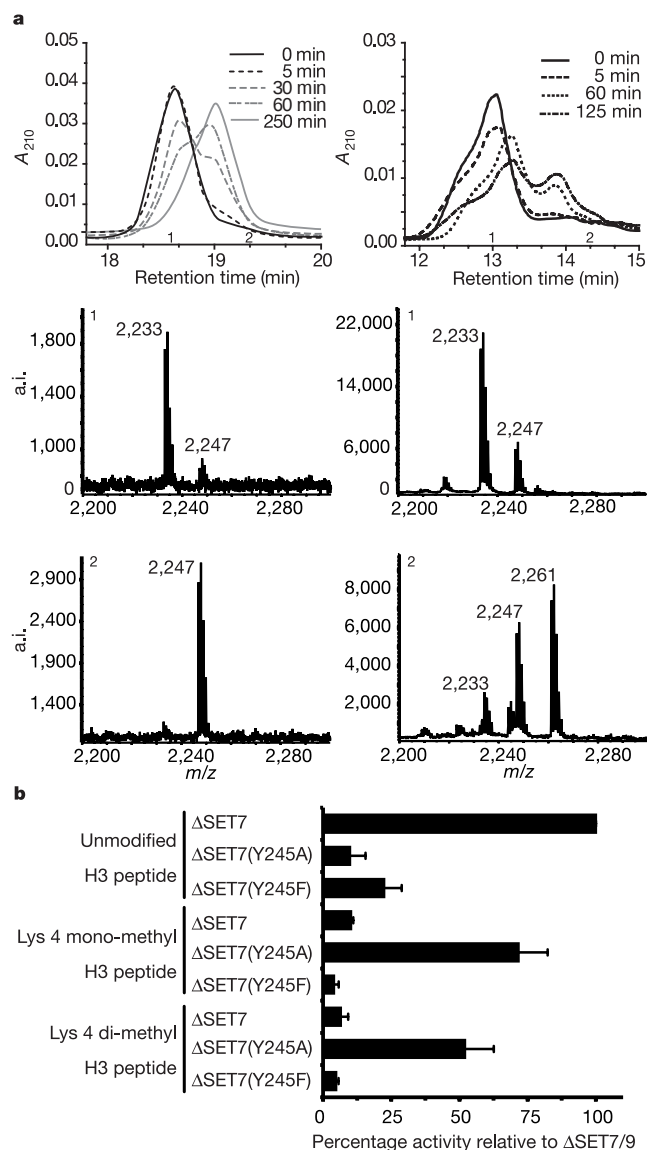


Figure 3 Activity of SET7/9. **a**, Analysis of methylation reaction products. The left-hand panels are for SET7/9; the right-hand panels are for G9a. The upper panel shows methylation of an H3 peptide at various time intervals as resolved by high-performance liquid chromatography (HPLC). The middle and lower panels show representative matrix-assisted laser desorption/ionization–time-of-flight (MALDI–TOF) analysis of fractions collected from the HPLC. Examples are from fraction points 1 and 2, respectively. A_{210} , absorbance at 210 nm; a.i., arbitrary intensity. **b**, Histone methyltransferase assay of SET7/9 and the point mutations Y245A and Y245F, with H3 peptide as substrate either unmodified, mono- or di-methylated at Lys 4.

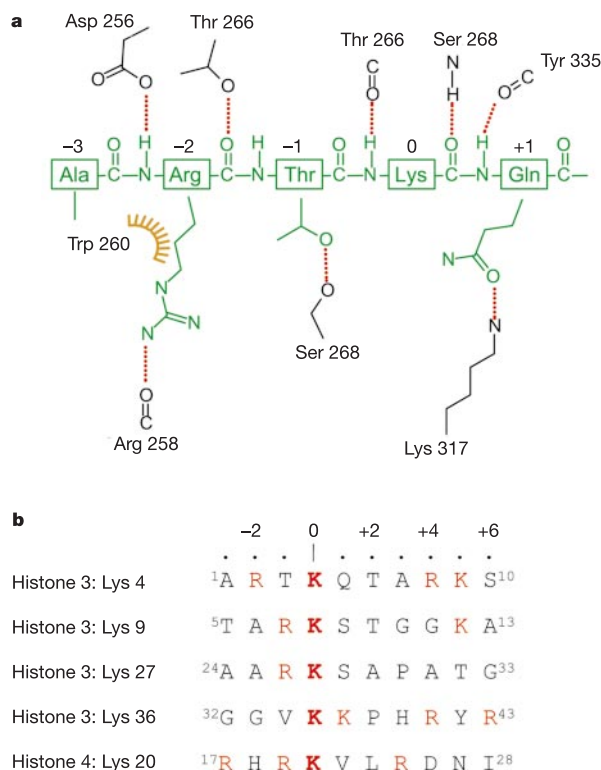


Figure 4 Peptide interactions with SET7/9. **a**, Schematic representation of interactions made by the histone H3 peptide in complex with SET7/9. The positions of residues beyond the +1 position are not shown because they are affected by lattice contacts. **b**, Sequences of histones subject to methylation, aligned according to the position of their target lysine residues. Basic residues are highlighted in red.

donates hydrogen bonds to both the invariant Tyr 245 and to a tightly bound water molecule (W1). We are confident that the lysine is acting as a hydrogen-bond donor in both cases because of the nature of the other hydrogen bonds made by Tyr 245 and W1 (Fig. 2a). The orientation of the lysine amine group is such that the amine-methyl bond is aligned towards the sulphur atom of the AdoHcy. Moreover, it is directed at sulphur along the tetrahedral vector corresponding to the (S)-location of the methyl group in AdoMet¹⁷. In other words, the geometry is exactly that expected for the trajectory of the methylation reaction and this feature will provide an important entropic contribution towards enzyme-mediated catalysis (Fig. 2c).

Our previous data¹¹ suggested that SET7/9 could not use mono-methyl Lys 4 peptide as a substrate for further methylation. Thus we have further analysed the methylation reaction. Reaction mixtures were set up with unmodified peptide and AdoMet, and conditions found where the reaction was driven essentially to completion. As shown in Fig. 3a, high-performance liquid chromatography (HPLC) separation of the reaction products resulted in a peak that eluted at a different position to the unmodified peptide, and matrix-assisted laser desorption/ionization (MALDI) analysis of this material revealed only mono-methylated peptide. Taken with our previous data, these results support the conclusion that SET7/9 catalyses the addition of a single methyl group to its target lysine. The reason for SET7/9 acting as a mono-methyltransferase is apparent from our crystal structure. The peptide used for crystallization was synthesized using mono-methylated lysine at position 4. In the crystal structure there is well-defined electron density for this methyl group in just one position. The arrangement of the two hydrogen-bond acceptor groups (for the lysine amine) provides an immediate explanation for the lack of rotation about the CE–NZ bond; Tyr 245 and W1 not only make favourable interactions that stabilize the observed rotamer, but they also preclude, on steric grounds, a methyl group in either of the two other positions that would orient a lone pair of electrons on the lysine nitrogen towards the sulphur of the AdoMet. Thus the structure shows that the arrangement of protein side chains and, indirectly, water molecules at the active site of SET7/9 is such that it can only catalyse the addition of a single methyl group to the lysine amine.

Inspection of the sequence of other SET proteins⁴ near the active site suggests why many of these enzymes catalyse di- or tri-methylation of their target lysines. Only Tyr 245 and Tyr 335 are considered invariant across the SET family, many other residues are highly variable. So, for example, Tyr 305 is substituted for a valine residue in SUV39H1 (a tri-methylase) and a proline in G9a (a di-methylase). Both of these substitutions would seem likely to produce a cavity at the active site that could accommodate a methyl substituent on the lysine amine. We have previously reported that the Y245A mutation in SET7/9 leads to a marked reduction in HMTase activity¹¹. We now show that although this mutant has greatly reduced activity with respect to an unmodified lysine substrate, it has substantial activity if assayed with mono- or with di-methylated Lys 4 substrate (Fig. 3b). Thus, Y245A is able to convert mono-methylated substrate to tri-methylated product. The structure readily provides a rationale for these observations. The chemistry of methyl transfer dictates that for di- and then for tri-substitution to occur, the existing methyl groups on the lysine would first have to be positioned approximately either at the OH of Tyr 245 or W1 sites, and finally in both of these positions. Removal of the aromatic ring of Tyr 245 creates a site for the first attached methyl group, thus enabling a second addition. The fact that the mutant enzyme can generate tri-methylated lysine further implies that the Y245A mutation also disrupts the local structure, so that W1 is displaced, allowing di-methylated lysine to bind appropriately for a third methylation.

The histone peptide binds in a largely extended conformation into a shallow groove, as shown in Fig. 1b. The binding is mediated

by a network of hydrogen bonds and salt bridges, involving both the main chain and side chains of the peptide (Fig. 4a). The target Lys 4_{PEP} residue is located approximately at the centre of the defined peptide, and the interactions of non-conserved SET7/9 residues with the peptide seems to account for the enzyme's specificity. Arg(–2)_{PEP} (residues are numbered relative to the target lysine) forms a hydrogen bond with the main-chain carbonyl of residue 258, such that the alkyl component of its side chain packs against the face of Trp 260. Thr(–1)_{PEP} forms a hydrogen bond with Ser 268, and Gln(+1)_{PEP} forms a hydrogen bond with the side chain of Lys 317. Arg(–2)_{PEP} therefore seems to have a particularly important role in mediating substrate specificity. Moreover, inspection of the various histone target sequences (Fig. 4b) shows that only the Lys 4 site of histone H3 contains a basic residue in the –2 position. The other histone target sites all have a basic residue either at the –1 or +1 position. Both activity measurements¹¹ and peptide affinity measurements, using a peptide in which Arg(–2)_{PEP} is mutated to an alanine (data not shown), support the hypothesis that this arginine residue contributes significantly to peptide binding to SET7/9. It seems that the structure of the ternary complex of SET7/9 presented here will provide a good model for assessing the probable substrate specificity of other SET proteins from their primary sequence. □

Methods

Protein constructs

ΔSET7/9 (residues 52–366) and G9a (621–1000) were expressed as glutathione S-transferase (GST) fusions in pGEX 6P1 in *Escherichia coli* BL21. The GST was removed by overnight treatment with PreScission Protease (Amersham) before gel filtration. Preparation of ΔSET7/9 in D₂O for NMR studies resulted in a series of N-terminal degradation products that remained catalytically active. Subsequently, ΔΔSET7/9 (108–366) was prepared as above and found to be stable for growth in D₂O, and was consequently used for further NMR and crystallography experiments. Site-directed alanine mutations were introduced using the Stratagene Quikchange mutagenesis kit, and mutations were confirmed by DNA sequencing and electrospray mass spectrometry. Synthetic peptides were prepared by W. Mawby, and AdoHcy was obtained from Fluka.

Crystallography

Protein stock solution was prepared at 100 mg ml^{–1} in 50 mM Tris, pH 7.0, 100 mM NaCl, and then incubated with a twofold molar excess of mono-methylated Lys 4 ten-residue peptide and AdoHcy. Crystals were grown by vapour diffusion at room temperature as hanging drops. Drops were prepared by mixing equal volumes of protein complex with reservoir solution containing 0.1 M Tris, pH 7.8, and 22% PEG 3350. Crystals were first transferred into mother liquor augmented with an additional 5% PEG 400, before plunging into liquid nitrogen. Data were collected from flash-cooled crystals at 100 K on a Raxis-II detector mounted on a Rigaku RU200 generator. Diffraction data were integrated and scaled using DENZO and SCALEPACK¹⁸. The structure was solved by molecular replacement using our previous model (1H31.brk) with AMORE. Subsequent refinement was carried out using REFMAC5¹⁹ and manual model building in O²⁰.

NMR

NMR spectra were recorded at 25 °C on Varian Inova spectrometers operating at ¹H frequencies of 600 MHz and 800 MHz. Protein samples (approximately 0.5 mM) were prepared in 50 mM Tris–HCl, 0.5 mM Tris(2-carboxyethyl)phosphine hydrochloride, 10% D₂O, pH 6.5.

HMTase activity measurements

The methyltransferase activity of SET7/9 and mutant constructs described in the text were determined in a reaction volume of 20 μl containing 3 μM AdoMet supplemented with [methyl-³H]AdoMet (4 μCi) (Amersham) and 750 nM purified methylase in reaction buffer (50 mM Tris, pH 8.5, 100 mM NaCl, 1 mM EDTA, 1 mM dithiothreitol (DTT)) with 50 μM histone peptide. After incubation at 37 °C for 60 min the reaction was vacuum-blotted onto membrane (Hybond-C; Amersham), washed, and the activity was measured by scintillation counting.

Analytical analysis of histone methylation

The histone methyltransferase assay was carried out at 37 °C in 50 mM Tris–HCl, pH 8.0, 100 mM NaCl, 1 mM DTT, with 300 μM AdoMet (Fluka), 100 μM H3 peptide (ARTKQTARKSTGGKAPRKQY), 1.5 μM enzyme. At the time intervals indicated, an aliquot of the reaction was removed and quenched in 8 M urea and acidified with glacial acetic acid. The reaction products were separated by reverse-phase HPLC (Jasco) on a Zorbax 300SB-C18 column (Rockland Technologies) using a gradient from 0% to 40% acetonitrile in the presence of 0.05% trifluoroacetic acid at 55 °C. Fractions from the peptide peak were analysed using a Reflex III MALDI–time-of-flight mass spectrometer (Bruker Daltonik) to obtain positive ion mass spectra.

Received 22 November; accepted 23 December 2002; doi:10.1038/nature01378.
Published online 22 January 2003.

1. Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389**, 349–352 (1997).

2. Turner, B. M. Histone acetylation and an epigenetic code. *Bioessays* **22**, 836–845 (2000).

3. Berger, S. L. Histone modifications in transcriptional regulation. *Curr. Opin. Genet. Dev.* **12**, 142–148 (2002).

4. Lachner, M. & Jenuwein, T. The many faces of histone lysine methylation. *Curr. Opin. Cell Biol.* **14**, 286–298 (2002).

5. Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41–45 (2000).

6. Zhang, X. *et al.* Structure of the neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase. *Cell* **111**, 117–127 (2002).

7. Spotswood, H. T. & Turner, B. M. An increasingly complex code. *J. Clin. Invest.* **110**, 577–582 (2002).

8. Feng, Q. *et al.* Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr. Biol.* **12**, 1052–1058 (2002).

9. van Leeuwen, F., Gafken, P. R. & Gottschling, D. E. Dtlp modulates silencing in yeast by methylation of the nucleosome core. *Cell* **109**, 745–756 (2002).

10. Jenuwein, T. Re-SET-ting heterochromatin by histone methyltransferases. *Trends Cell Biol.* **11**, 266–273 (2001).

11. Wilson, J. R. *et al.* Crystal structure and functional analysis of the histone methyltransferase SET7/9. *Cell* **111**, 105–115 (2002).

12. Jacobs, S. A. *et al.* The active site of the SET domain is constructed on a knot. *Nature Struct. Biol.* **9**, 833–838 (2002).

13. Trievel, R. C., Beach, B. M., Dirk, L. M. A., Houtz, R. L. & Hurley, J. H. Structure and catalytic mechanism of a SET domain protein methyltransferase. *Cell* **111**, 91–103 (2002).

14. Min, J., Zhang, X., Cheng, X., Grewal, S. I. & Xu, R. M. Structure of the SET domain histone lysine methyltransferase Clr4. *Nature Struct. Biol.* **9**, 828–832 (2002).

15. Santos-Rosa, H. *et al.* Active genes are tri-methylated at K4 of histone H3. *Nature* **419**, 407–411 (2002).

16. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).

17. Hofmann, J. L. Chromatographic analysis of the chiral and covalent instability of S-adenosyl-L-methionine. *Biochemistry* **25**, 4444–4449 (1986).

18. Otwinowski, Z. & Minor, W. in *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 556–562 (SERC Daresbury Laboratory, Warrington, 1993).

19. CCP4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).

20. Jones, T. A., Zhou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).

Supplementary Information accompanies the paper on *Nature's* website
(<http://www.nature.com/nature>).

Acknowledgements We are grateful to G. Dodson and S. Smerdon for critical reading of the manuscript, and to Y. Shinkai and Y. Tanaka for the gift of the G9a clone. NMR spectra were recorded at the MRC Biomedical NMR Centre.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.J.G.
(e-mail: sgambli@nimr.mrc.ac.uk). Coordinates for the SET7/9 ternary complex have been deposited with the Protein Data Bank under accession code 1o9s.

corrigendum

Contemporary fisherian life-history evolution in small salmonid populations

Mikko T. Koskinen, Thron O. Haugen & Craig R. Primmer

Nature **419**, 826–830 (2002).

The neutrality test equation applied in this Letter ($F = (N_e \sigma_{GB}^2) / (h^2 \sigma_{GW}^2 t)$) was incorrect: the correct equation is $F = (N_e \sigma_{GB}^2) / (\sigma_{GW}^2 t)$. Consequently, the numbers in the right three columns of the original Table 1 are wrong (corrected below). In addition, some variance estimates in Table 1 and some values of Fig. 2 were reported incorrectly (the entire correct Table 1 and the correct Fig. 2 are available from the authors). These errors do not affect our conclusion that the populations evolved predominantly as a result of natural selection. We thank W. G. Hill for bringing these errors to our attention. □

Table 1				
Trait	Pairwise comparison	$F_{1, \infty}$	P	N_e (sign)
Length at termination (mm)	Les-Ht	37.7	***	2.5
	Les-Aur	0.78	0.38	272
	Ht-Aur	8.63	**	17.7
Yolk-sac volume (mm ³)	Les-Ht	29.0	***	3.21
	Les-Aur	59.2	***	3.59
	Ht-Aur	3.17	0.07	48.3
Growth rate (mm/ΔD)	Les-Ht	20.5	***	4.55
	Les-Aur	6.31	**	33.7
	Ht-Aur	41.5	***	3.68
Survival (%)	Les-Ht	5.90	*	15.8
	Les-Aur	3.13	0.08	68.0
	Ht-Aur	1.20	0.27	128
Incubation time (days)	Les-Ht	4.02	*	23.1
	Les-Aur	1.56×10^{-7}	>0.99	1.36×10^9
	Ht-Aur	7.82	**	19.5
Swim-up length (mm)	Les-Ht	0.26	0.61	363
	Les-Aur	0.04	0.84	5.79×10^3
	Ht-Aur	2.73×10^{-8}	>0.99	5.60×10^9
Hatching length (mm)	Les-Ht	1.46×10^{-8}	>0.99	6.38×10^9
	Les-Aur	0.78	0.38	274
	Ht-Aur	2.16×10^{-8}	>0.99	7.07×10^9

DNA on the slide

Microarray readers, tissue microarrays and high-throughput screening.

Immobilizer

Flowgen www.flowgen.co.uk

Spot the difference

The Immobilizer is an open, 'read-to-spot' polymer microarray slide with a transparent spotting area made of anthraquinone photo probes to enable one-step coupling (spotting) of amino-linked DNA/RNA, -LNA or -PCR. Between the spots there is low, nonspecific binding of DNA and fluorochrome labelled proteins. The device eliminates the need for activation steps, blocking, baking or UV cross-linking. Applications include clone mapping, linkage analysis, highly multiplexed mutation detection, gene expression monitoring and high-throughput screening.

DNAScope AT

**Biomedical Photometrics Inc./
Clondia Gmbh** www.genefocus.com

All in order

Up to six Clondia ArrayTubes can be read automatically in sequence with the DNAScope confocal fluorescence reader. It contains two lasers housed in a small case that sits on the laboratory bench. Reading time is less than a minute per tube. The ArrayTube system allows direct implementation of high-quality DNA arrays into a standard micro-reaction tube.

Sirius HT and Discovery HT-R

MWG Biotech www.THE-MWG.com

Reading aids

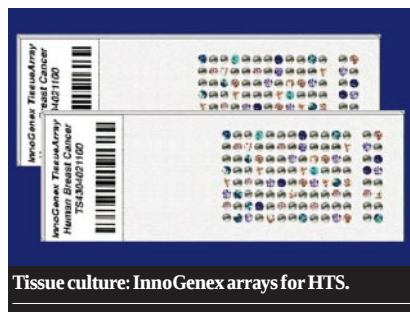
MWG has introduced two microplate readers, each with KC4 data analysis and reduction software making them suitable for integration with laboratory automation systems. The dual optics Sirius HT microplate reader features fluorescence/time-resolved fluorescence/luminescence readings as well as absorption readings without filters. The Discovery HT-R with monochromator is an eight-channel spectrophotometer with a reading time of 5 seconds per 96-well plate and 11 seconds per 384-well plate.

Tissue microarrays

InnoGenex www.innogenex.com

For high-throughput screening

InnoGenex supplies tissue microarrays for high-throughput screening of *in situ* gene expression in target validation studies. Microarrays from diseased and normal human tissues, mouse normal tissues and rat normal tissues are available. The histologically characterized slides can be used to deter-



mine the correlation of the gene and protein expression profile in normal and diseased tissues. Low-density arrays contain 20–50 elements per slide, while the high-density arrays have 100–200 elements per slide. Applications include rapid screening of differential expression of genes, identification of novel antibody markers and animal model studies.

Universal Reference RNA

BD Biosciences www.bdbiosciences.com

Control for microarray experiments

Universal Reference RNA from BD Biosciences is derived from whole-tissue sources and made by pooling the total RNA extracts from a collection of different human tissues, providing coverage of more than 92% of genes. The RNA can be used to compare microarray data that are generated on multiple days.

HybChamber Mica

GeneMachines www.genemachines.com

Under cover of darkness

The HybChamber Mica provides protection from ambient light degradation of fluorescent labels during hybridization. The darker, semi-transparent top plate also allows convenient visualization of microarray slides during all hybridization steps. The device holds two slides, has butterfly screws to allow easy tightening and loosening of the chambers, and can be used in temperatures of up to 70 °C.

PathwayFinder GEMarrays

SuperArray Bioscience www.superarray.com

Guess not

This system groups pathway representative markers in a cDNA array. Because the expression of marker genes is associated with the activation/inactivation of well defined biological pathways, a rapid assessment of many pathways can be made in one hybridization experiment. Three PathwayFinder GEMarrays are available, for signal transduction, cancer

and cAMP/Ca²⁺. Once the pathways have been identified, pathway-specific GEMarrays can be used for characterization of individual signal transduction pathways of NfκB, MPK, NFAT, TGFβ, p53, G-protein coupled receptor, cAMP/Ca²⁺, insulin, androgen and oestrogen signalling. The company also offers angiogenesis, metastasis, cell cycle, apoptosis, DNA damage and repair arrays for studies of cancer.

These notes are compiled in the Nature office from information provided by the manufacturers.

Automation on the move

The insatiable demands of the pharmaceutical industry are helping to drive labs into the automated world of industrial production lines. Tim Chapman joins the quest for speed.

Since clinical laboratories began using robotics in the early 1980s to manage their thousands of samples a day, automated systems have become commonplace. By rapidly and tirelessly performing processes that would have been done manually — from sample storage and retrieval to running customized assays — modern automated systems can greatly increase the throughput of a laboratory, free up researchers from repetitive tasks, and monitor and manage the raw data produced.

Automation is now increasingly common in agricultural, food and environmental analytical labs, but it is the demands of the pharmaceutical industry for high-throughput screening of potential drug candidates (see *Nature* **418**, 453–459; 2002) that is most influential in driving the technology forward. Researchers now have whole genomes of potential drug targets at their disposal for testing against an array of drug candidates, and high-throughput automation is rapidly becoming a necessity.

“Automation is used in a broad range of approaches in pharmaceuticals, from early combinatorial creation to high-throughput

screening,” says Robin Felder, director of the Medical Automation Research Center at the University of Virginia in Charlottesville and president of the Association for Laboratory Automation. “The skills of creating a new drug start with synthesizing novel molecules — it used to be one scientist, one molecule, one week. Now, one molecular chemist can turn out 100 good compounds a week.”

The fast track

Many pharmaceutical and biotechnology companies would be unable to function without high levels of automation. A typical end-user is Cypotex, based in Macclesfield, UK, which aims to improve the efficiency of drug discovery by using industrial-scale testing and proprietary software to predict how drug candidates will act in the body.

The assays that could rule out an otherwise promising drug candidate are for absorption, distribution, metabolism and excretion. To avoid a bottleneck in the drug-discovery process, these assays have to be carried out as rapidly as the drug candidates are identified. “We have automated the assays, which are currently



David Leahy: screening can benefit greatly from automation.

running at 100–200 compounds a week in major pharmaceutical companies, to where we can run 2,000–5,000 compounds a week,” says David Leahy, founder and chief scientific officer at Cypotex. “We think there’s a potential for running, if we pick one assay, say 750,000 compounds a year.”

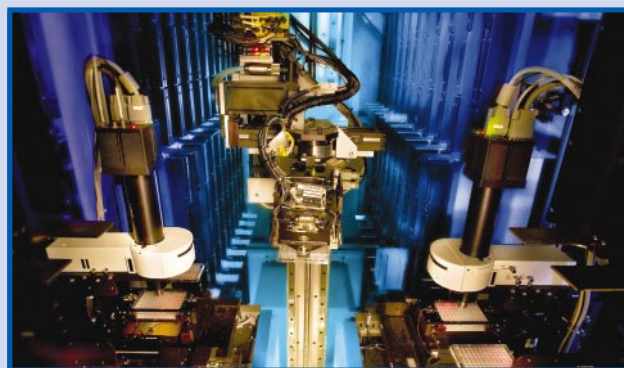
A STRUCTURED APPROACH

Mapping the detailed atomic structure of a target protein is vital for identifying drug candidates that can bind with it. Elucidating protein structure by crystallography has for years been a laborious operation carried out by specialists, but automation is now beginning to bring the process up to the speeds demanded by high-throughput systems.

Automated crystallography systems for the general research market have been available for some time from specialists such as Tecan in Männedorf, Switzerland, and BSI Proteomics of Gaithersburg, Maryland. But dedicated drug-discovery companies are now forcing the pace, with the vast majority of commercial protein crystallography done by just two companies, Syrrx and Structural GenomiX, both based in San Diego.

Syrrx wants to automate every stage of protein-structure determination. It deploys a family of whimsically named robots, beginning with a high-throughput robot called Sonic Hedgehog that can purify about 100 proteins in five hours. The proteins are then crystallized by Agincourt, a robot named for English king Henry V’s application of superior technology to overcome extreme odds. Agincourt crystallizes proteins from droplets of just 50 nanolitres, about a fortieth of the volume of previous systems. “Everything is miniaturized,” says business development director Ned David. “These droplets enable our whole approach to work in a way that is economically feasible. You can’t do nano-drops without automation because you can’t set your pipette for 50 nanolitres.”

To match the demands of Agincourt, Syrrx has two imaging-station robots called Fort Knox and Fort Bliss, each capable of scanning over a



Syrrx’s high-throughput image-scanning robot.

million images a day for suitable protein crystals. These are then taken to Syrrx’s dedicated synchrotron beamline at the Advanced Light Source in Berkeley for X-ray diffraction. Data are collected by another robot called Robohutch, which David says can increase throughput at this stage tenfold.

There is still scope for further automation, particularly in the first stages of protein production. Syrrx is investigating how more mammalian proteins can be made in eukaryotic cell systems, rather than from genes cloned in bacteria. **T.C.**

AARON FELDMAN

But this level of automation brought technical challenges. Specialist consultancy The Automation Partnership (TAP) in Royston, UK, worked with Cyprotex to adapt TAP's proprietary BasePlate liquid-handling robots to Cyprotex's particular requirements. Cyprotex also automated many of its analytical instruments to run standardized tests on compounds coming through the high-throughput process.

Cyprotex is now in a position to extend its automated systems to carry out a wider range of important assays. "We think automation gives us a screening capacity that is at least an order of magnitude higher than a typical pharmaceutical company, at a unit cost that is possibly 20-fold less," says Leahy.

Small is beautiful

Benchtop automated liquid handling and sample-dispensing systems are now routine in most life-science laboratories. Such systems will become even more ubiquitous with the introduction of a new wave of lower-cost modular devices with much the same functionality as the systems used by big-pharma labs. "The pharma labs have the deepest pockets, but a new generation of automated systems is becoming available that's more the 'Volkswagen' of the market," says Felder.

Xiril, a young Swiss company founded in Hombrechtikon by a team of specialists in robotic liquid-handling, has a range of pipetting robots that costs about 20–30%

less than comparable systems. The Xiril robots are primarily aimed at low-to-medium-throughput operations, particularly for processes that have not yet been automated.

"We believe we can fill a market for liquid-handling processes that are often done manually, even in labs that are highly automated," says the firm's chief executive Heinz Abplanalp. Xiril has also developed automated systems for magnetic-bead separation.

At the more expensive end of the liquid-handling market, the trend is towards increasing miniaturization. Given the high cost of many reagents used in biomedical experimentation, considerable savings can be made simply by using less. Allegro Technologies, a Dublin-based company spun out of Trinity College, has the proprietary technology to dispense droplets up to 1,000 times smaller than those from standard equipment. The company claims that this could reduce the cost of some experiments sixfold.

Allegro's latest range of pipetting systems uses electromagnetically controlled valves to deliver droplets of 20 microlitres down to 20 nanolitres in volume. The firm has licensed its technology to US instrument manufacturers Gilson of Middleton, Wisconsin, and Beckman Coulter in Fullerton, California, and has also put the pipettes into its own robotic system.

Being able to manipulate tiny volumes

of liquid means that entire lab processes can be miniaturized. Using technology developed for the semiconductor industry, chips a few centimetres square can be etched with microscopic channels in which routine preparation procedures and assays can be carried out.

Caliper Technologies in Mountain View, California, was the first to market these 'lab-on-a-chip' devices. Its 250 HTS system can use a range of chips for enzymatic or cell-based drug screens; some assays require fewer than 500 cells.

Caliper's chips use small electrical currents or vacuum-based pressure techniques to move fluids through the channels. Gyros in Uppsala, Sweden, is developing a range of microlabs for processes such as protein purification in the form of compact discs. A robotic workstation loads the CD with reagents and spins it at precisely controlled speeds, driving the fluids through the channels by centrifugal acceleration. Gyros's first product, the Gyrolab MALDI, can prepare up to 96 protein samples simultaneously for mass spectrometry. Tecan has also launched a similar product — its LabCD devices for pharmacological assays.

Action stations

Workstations, stand-alone automated solutions for specific lab tasks such as DNA sequencing or immunoassay tests, continue to be dominated by established

CULTURAL REVOLUTION

Cell culture, either for the cells themselves or their products, is traditionally a manual process that demands hours of repetitive, painstaking work to ensure absolute sterility under exacting conditions. Growing a mammalian cell culture for use in a screen for drug activity or toxicity typically takes 48 hours, so almost half the working week is gone before screening can begin.

But with automated cell-culture systems, screening could start on Monday morning on cultures left to grow over the weekend. Robotic systems can also provide a consistency of procedure and sterility that the best technicians cannot match, ensuring less variability between batches.

The Automation Partnership (TAP) in Royston, UK, introduced its pioneering Cellmate workstation in the late 1980s for the relatively large-scale production of therapeutic reagents and vaccines. This has now been joined by a new system, SelecT, developed in association with a consortium of six major drug companies, including GlaxoSmithKline and Pfizer.

SelecT is aimed at labs that need an automated system for processing up to 168 different mammalian cell lines simultaneously, making it ideal for multiple batches. The system can automatically maintain, expand, process and harvest multiple cell lines, assess their viability, and distribute them onto a maximum of 300 microtitre plates. "Cellmate is not so good at dealing with just-in-time production of ready-to-go microplates," says Mark Beggs, head of consulting at TPA. "SelecT can handle all the capabilities you need to take cells through to incubation to microplates."

Other commercial systems provide an even greater level of integration. Cytogration, a subsidiary of Brandel in Rockville, Maryland, has developed a

robotic system that performs both cell culture and screening assays for drug candidates. The standard system handles up to 504 plates, and automates cell production, preparation of membrane-bound cells and *in vitro* screening.

An automated system has primarily to maintain a stable environment and reduce the risk of contamination, says Sean Sales, applications consultant at RTS Life Sciences in Manchester, UK. The system also needs to be dynamic in terms of knowing everything, so that the robot itself isn't slowing things down, he says. RTS has developed a flexible cell-culture system called

acCellerator which has a very simple liquid-handling method and three parallel robot arms to process the flasks and plates in parallel. "It has a number of peristaltic pump units that allow a very short liquid line for reagents, so you don't have a lot of tubing to change over between cell lines," says Sales.

T.C.



The acCellerator: flexible cell culture from RTS.

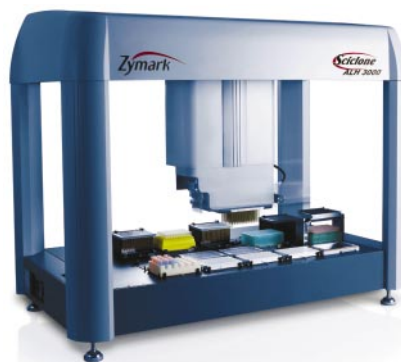
names. The Biomek FX liquid-handling workstation from Beckman Coulter, used extensively in DNA preparation, is a sophisticated descendant of the Biomek 1000 system introduced in 1986. The latest model can simultaneously automate three liquid handlers on the same unit.

The popular BioRobot series of automated workstations made by QIAGEN Instruments near Zurich, Switzerland, offers a range of dedicated platforms for high-throughput genomics. The most recent entrant, the BioRobot Protein LS system, integrates all the software and hardware needed to purify and prepare proteins for study using X-ray crystallography or nuclear magnetic resonance.

Zymark, based in Hopkinton, Massachusetts, continues to develop its Staccato range of liquid-handling workstations, based around the Sciclone liquid-handling robots. The Sciclone ALH3000, launched in autumn last year, can dispense into microplates of up to 1,536 wells, with a new head design that can be rapidly switched to pipette volumes of 1 millilitre down to 100 nanolitres.

"The general trend we are seeing is to integrate a greater number of functions into a single system that can carry out a related set of applications," says Mark Roskey, vice-president of marketing at Zymark.

Advances in detection technology are also boosting the capabilities of workstations. The plate::screen workstation



Zymark's latest Sciclone robot brings added flexibility to liquid handling.

for microplate processing, launched last year by Carl Zeiss of Jena, Germany, incorporates the first commercially available multimode detection system to deliver high-performance data from microplates of up to 1,536 wells. The plate::vision detector reads 96 wells in parallel with three detection modes — absorbance, fluorescence and luminescence — allowing it to analyse a 1,536-well plate in 20 seconds.

All together now

In theory, high-throughput systems should be able to screen about 100,000 compounds a day against a single disease target. But such rates can rarely be maintained consistently because of problems and delays in

transferring sample plates between individual tests and preparation processes. Drug companies typically screen about 25,000 compounds a week, says Mark Beggs, head of consulting at TAP. "What that's really telling you is that coordination of subordinate processes is less than optimal."

Integrating the various stages of automation into a laboratory can mean that researchers have to adapt to a more industrial production-line environment. "The hard changes are organizational ones, and getting people away from a research-driven culture to a production-driven culture," says Beggs. That requires attention to bottlenecks and the time taken to transfer plates between individual stages rather than just buying a faster robot, he says. "While some companies get faster by buying the latest thing from the box, more are increasingly becoming faster through process-time reduction."

In addition, commercial high-throughput screening demands extreme flexibility and robustness — a typical screening department in a major drug company will run 20–50 new assays a year against a library of up to 1 million compounds, with each assay involving up to 100 separate activities. Downtime for reconfiguration or replacement of an automated system can significantly reduce effective throughput. Research by automated solutions provider RTS Life Sciences of Manchester, UK, suggests that

SAMPLED DELIGHTS

One of the remaining challenges for high-throughput automation comes at the beginning of many processes — the storage and retrieval of samples. Keeping track of samples can be a major headache for many labs, with researchers often spending hours locating a single specimen.

A pioneer in this field is Biophile, based in Charlottesville, Virginia. Biophile has developed a fully automated storage and retrieval system, which operates at temperatures down to the -80°C required for storing materials such as RNA and tissue samples. The unit has built-in robotics that pick from a rotating vertical stack of almost 1,000 microplates or racks of cryogenic vials.

This month, the company is launching its individual vial retriever, which can automatically re-rack, scan, weigh and sort vials. The retriever will have a capacity of over 22,000 cryovials, which it can select and deliver at an average rate of 200 per hour. "This could double the productivity and rate of working with samples," says Sean Graves, the company's vice-president for technology.

The storage system is designed to be compatible with a wide range of equipment, and the vial retriever is designed to work with LIMS and to take samples directly to workstations by a conveyor system. As with LIMS, the demand for automated storage systems will be increased by regulatory pressure for drug companies and research labs to maintain the security of medical specimens for long periods of time.

"Automated storage and retrieval systems is the clear application that's going to surprise everyone," says Robin Felder, president of the Association for Laboratory Automation. "It's going to surprise everyone because you're going to have to have this."

T.C.



Boxing clever: Biophile's storage and retrieval system.

each high-throughput screening system typically has an instrument replaced every six months because of the continual advances in instrument technology, meaning that research is frequently put on hold while the system is upgraded.

Some parts of common high-throughput screening processes have proved harder to automate than relatively simple tasks such as sample dispensing. Protein crystallography and cell culture, for example, are now opening up to full automation (see 'A structured approach', page 661 and 'Cultural revolution', page 663), whereas others, such as high-throughput mass spectrometry, remain more intractable.

Some automation providers have improved the overall throughput of a complex process by creating a series of specialized platforms that can be fully integrated to track each sample and its accompanying data through the whole process. The BioCube platforms from Protodyne in Windsor, Connecticut, are designed to address the worst bottlenecks in biotech processes. The range so far tackles liquid handling, the polymerase chain reaction, colony and plaque picking, and gel electrophoresis. The most recent addition is the SNPCube, which has a daily identification rate of up to 90,000 single-nucleotide polymorphisms — the single base differences between individuals that are sometimes linked to disease.

"The BioCube can become a complete system where data and processes are traced from the second the first plate enters the system," says European product manager Volker Knack. Eight- to tenfold increases in throughput can be achieved across the entire drug-discovery process, Knack says.

Organization

The development of automated systems in the life-science laboratory has gone hand-in-hand with the information-management systems needed to handle the masses of data generated. But the conventional laboratory information management systems (LIMS) developed for analytical chemistry operations are generally too inflexible for bioscience work. As well as producing a greater volume and complexity of data, bioscience researchers usually need a wider view of the relationships between data sets.

This has led to a new generation of systems being designed by a range of firms from specialized software developers such as L.I.M.S. in Hollywood, Florida, with its StarLIMS package, to IBM Life Sciences with its DiscoveryLink data-integration system.

"LIMS holds it all together," says Felder. "The LIMS area is actually evolving itself — it used to be gathering data and organizing it, but it's now getting to be integrated so



Protodyne's BioCubes are aimed at easing bottlenecks in lab processes.

you can be at a single workstation and view what everyone is doing."

The newest entrant to this market is the Sierra Laboratory Workflow System range from Amersham Biosciences in Piscataway, New Jersey, which is aimed at high-throughput genomics research. As laboratories scale up, the amount of data they generate increases hugely, as does the challenge of extracting useful information from the data (see *Nature* **419**, 751–757; 2002). Sierra is designed to mirror the natural workflow of a lab, linking manual and automated processes into one integrated management system.

The Sierra system interfaces closely with the software on dedicated automated equipment such as Amersham's MegaBACE DNA sequencer, so that additional equipment can be integrated into the system. But equipment from other suppliers may not be compatible, either because individual instruments don't provide a suitable interface or because their controlling software cannot be programmed to talk to other systems.

"At this stage of development, most LIMS cannot fully drive a whole lab process, so there's still an amount of human involvement," says John Schneider, vice-president for marketing at Amersham's informatics division.

Integrating LIMS with a range of automated platforms can demand considerable programming skills. The Nautilus system developed by Thermo LabSystems, based in Altrincham, UK, is designed to simplify the process by converting data from instruments and other applications into a common XML-based file format. "This means it can handle the various robotics in a lab that are constantly

changing," says marketing manager Adrian Fergus. "What's important is to provide open hooks from the LIMS to allow it to quickly and easily integrate with new robotic equipment and new applications."

Such flexible LIMS are likely to become a necessity for many drug labs because of the ruling by the US Food and Drug Administration that data on drug development must be kept in a format that will be accessible for evaluation no matter how data-storage technology evolves. Although this legislation has been in force since 1997, enforced compliance is only now being phased in, prompting a new wave of compliant systems. For example, Hamilton, a developer of fluid-handling systems in Reno, Nevada, has recently launched a new version of the controlling software for its Microlab robotic instruments.

As well as being sold as a 'ready-to-run' software package, Nautilus is also available as an application service, where labs pay a monthly fee for a hosted LIMS service delivered through the Internet. "It generally appears to be smaller labs that have shown an interest, where they do not have the IT resources in-house to manage a LIMS and are willing to have that system hosted," Fergus says. "What it avoids is the initial upfront cost of the licences and so spreads the cost more evenly and is more manageable."

As with all areas of laboratory automation, the aim of LIMS is to increase productivity and efficiency, says Felder. "The motivation in this industry is not eliminating jobs, it is getting products to market faster. This is one case where automation will create jobs rather than eliminating them," he says.

Tim Chapman is a freelance journalist based in Halifax, UK.

Company	Products/activity	Location	URL
Benchtop systems			
Advion Biosciences	Fully automated nanoelectrospray for mass spectrometry	Ithaca, New York	www.advion.com
Allegro Technologies	Nanovolume pipetting systems	Dublin, Eire	www.allegro-technologies.com
BD Biosciences	Multiwell AutoSampler for flow cytometers; high-throughput solubility scanner	Oxford, UK	www.bdeurope.com
BioRobotics	MicroGrid automated arrayer and BioPick automated colony picker	Cambridge, UK	www.biorobotics.com
Bio-Tek	Automated microplate handling and reading instrumentation, spectrophotometers	Winooski, Vermont	www.biotek.com
BMG Labtechnologies	Microplate readers and handling systems	Offenburg, Germany	www.bmglabtech.com
Boston Innovation	Automated sample delivery	Cambridge, Massachusetts	www.bostoninnovation.com
Colibri Robotics	Suppliers of Xiril pipetting robots and Plato automated microplate processors	New Castle, Delaware	www.colibrirobotics.com
Genevac	Automated solvent removal systems	Ipswich, UK	www.genevac.com
GenoVision	GenoM-6 automated system for DNA isolation and purification	Westchester, Pennsylvania	www.genovision.com
Gilson	Automated liquid-handling, pipetting and protein crystallization instruments	Middleton, Wisconsin	www.gilson.com
IGEN International	Automated assay analysers using electrochemiluminescence detection	Gaithersburg, Maryland	www.igen.com/home.htm
MJ Research	DNA Engine thermal cyclers, 96- and 384-well hard-shell plates	Waltham, Massachusetts	www.mjrc.com
OptiCell	Automated cell-culture devices	Westerville, Ohio	www.opticell.com
Titertek	Precision liquid-handling and bench-top automation of microplate assays	Huntsville, Alabama	www.titertek.com
Torcon Instruments	Precision fluid dispensing and automated instrumentation for microplate handling	Torrance, California	www.torconinstruments.com
Xiril	Pipetting robots	Hombrechtikon, Switzerland	www.xiril.com
Zinsser Analytic	Pipetting, extraction and liquid-handling robots	Frankfurt, Germany	www.zinsser-analytic.com
Microfluidics			
Aclara BioSciences	LabCard microfluidics arrays for high-throughput drug screening, multiplexed gene-expression analysis, and multiplexed SNP genotyping	Mountain View, California	www.aclara.com
Caliper Technologies	LabChip microfluidic systems for high-throughput screening for drug discovery, biological and genetic research	Mountain View, California	www.calipertech.com
Gyros	CD microlaboratory for sample preparation for protein mass spectrometry	Uppsala, Sweden	www.gyros.com
Workstations			
Applied Biosystems	DNA sequencers, mass spectrometers, workstations and automated instruments	Foster City, California	home.appliedbiosystems.com
Affymetric	GeneChip microarray systems	Santa Clara, California	www.affymetric.com
Agilent	Workstation for RNA, DNA or protein analysis, instruments for genomics and proteomics	Palo Alto, California	www.agilent.com
Applied Scientific Instrumentation	Automated systems for microscopy and fluorescence screening	Eugene, Oregon	www.asiimaging.com
AutoGen	Automated nucleic acid extraction instruments and colony/plaque-picking systems	Holliston, Massachusetts	www.autogen.com
Beckman Coulter	Automation tools from modular liquid-handling systems to integrated robotic systems	Fullerton, California	www.beckman.com
Biotope	Horizon range of high-performance FLASH chromatography systems	Charlottesville, Virginia	www.biotope.com
BioTec	Liquid-handling lab instruments, pipetting workstation, dispenser, microplate washer	Tokyo Japan	www.biotech.co.jp
Brandel	Automated sample preparation and harvesting, superfusion systems, microdispensers	Gaithersburg, Maryland	www.brandel.com
Bruker	Automated platforms for mass spectrometry, NMR, EPR and magnetic resonance	The Woodlands, Texas	www.bruker.com
Carl Zeiss	plate:explorer and plate::screen automation for ultra-high throughput screening	Jena, Germany	www.zeiss.de
Cell Robotics International	Microscope workstations for biomedical research	Albuquerque, New Mexico	www.cellrobotics.com
Chemical Diversity Labs	CombiSyn reactors for solid- and solution-phase parallel synthesis	San Diego, California	www.chemdiv.com
Cytogation	Automated high-throughput cell culture systems for <i>in vitro</i> screening	Rockville, Maryland	www.cytogation.com
deCODE Genetics	Emerald Biostructures products and robotic platform for protein crystallization	Bainbridge Island, Washington	www.decode.com
Dynex Technologies	Absorbance and luminescence microplate readers, microplate washers and automated workstations	Chantilly, Virginia	www.dynextechnologies.com
GeneMachines	Automated instruments for sample preparation, microarraying, DNA purification, oligonucleotide synthesis and colony/plaque picking	San Carlos, California	www.genemachines.com
Genetix	Picking, arraying, gridding, replicating, rearraying and microarraying robots	New Milton, UK	www.genetix.co.uk
Genomic Solutions	Automated instrumentation, software, consumables and services	Ann Arbor, Michigan	www.genomicsolutions.com
Gentra Systems	Automated equipment for nucleic acid purification	Minneapolis, Minnesota	www.gentra.com
Hamilton	Microlab automated liquid-handling workstations	Bonaduz, Switzerland	www.hamiltoncomp.com
Labcyte	Automated pipetting workstations, microplate washers, magnetic-bead washers and DNA-extraction working decks	Union City, California	www.labcyte.com
Lab Services	Automated liquid-handling instruments, fluorescence readers and microplate labs	Breda, the Netherlands	www.lab-services.nl
Liconic Instruments	Lab automation, automated incubators and storage systems	Mauren, Liechtenstein	www.liconic.com
Magnetic Biosolutions	Robotic workstations for automated biomagnetic separation	Stockholm, Sweden	www.magbio.com
MWG-Biotech	Thermal cyclers, automated systems for high-throughput sequencing and microarray production	Ebersberg, Germany	www.mwg-biotech.com
Nanogen	NanoChip automated workstation for SNP and STR analysis	San Diego, California	www.nanogen.com
QIAGEN	BioRobot workstation; automated nucleic acid, protein and cell purification	Venlo, The Netherlands	www.qiagen.com
Relab	Arrays, scanners, workstations and dispensers for biochips	Recklinghausen, Germany	www.relab.de
Sias	Xantus robotic liquid-handling platform; lxion robot-friendly microplate centrifuge	Hombrechtikon, Switzerland	www.sias.biz
Tecan	Laboratory automation for the life sciences, LabCD 'lab-on-a-disc' assays	Mannedorf, Switzerland	www.tecan.com
Tomtec	Liquid-handling and solid-phase extraction workstations; automated instruments	Hamden, Connecticut	www.tomtec.com
Union Biometrica	COPAS platforms for automated high-throughput analysis, sorting and dispensing of small multicellular organisms	Somerville, Massachusetts	www.unionbio.com
Velocity11	BioCel automation platform for high-throughput microplate processing, microplate dispensers and pipetting stations	Palo Alto, California	www.velocity11.com
Zymark	Sciclone liquid-handling stations, Staccato high-throughput workstations	Hopkinton, Massachusetts	www.zymark.com

Company	Products/activity	Location	URL
Integrated laboratory automation			
AB Controls	Custom hardware and software for laboratory automation and robotics	Irvine, California	www.abcontrols.com
accelab	Customized laboratory automation	Kusterdingen, Germany	www.accelab.de
Biometra	TRobot thermocycler for integration with robotic environments	Göttingen, Germany	www.biometra.de
Cellomics	ArrayScan HCS and KineticScan HCS for automated cell-based screening	Pittsburgh, Pennsylvania	www.cellomics.com
CyBio	Pipetting, liquid handling, incubation, and imaging systems for automated screening	Jena, Germany	www.cybio-ag.com
DataCentric Automation	Automated systems for protein crystallography	Nashville, Tennessee	www.titanceg.com/divisions
H+P Labortechnik	VARIOMAG magnetic stirrers, shakers and reaction blocks	Oberschleißheim, Germany	www.hp-lab.de
Hettich-Zentrifugen	ROTANTA 46 RSC robotic centrifuge	Tuttlingen, Germany	www.hettichlab.com
Hitachi High-Technologies	Automated instrumentation for mass spectroscopy, liquid chromatography, high-throughput purification, amino-acid analysis and spectroscopy	Tokyo, Japan	www.hitachi-hitec.com
IBM Corporation Life Sciences	DiscoveryLink platform for database integration; data-management systems	Somers, New York	www.ibm.com/solutions/lifesciences
J-KEM Scientific	Customized automated laboratory equipment	St Louis, Missouri	www.jkem.com/index.htm
KBiosystems	Robotic instrumentation for high-throughput genomics and proteomics	Basildon, UK	www.kbiosystems.com
Kendro Laboratory Products	Cyomat robot-accessible incubators	Hanau, Germany	www.kendro.com
Molecular Devices	Liquid-handling and microplate-processing; imaging; assay systems	Sunnyvale, California	www.moleculardevices.com
Process Analysis and Automation	Automated laboratory instruments and software	Farnborough, UK	www.paa.co.uk
Proteodyne	BioCube automated systems; LIMS middleware and integration software	Windsor, Connecticut	www.proteodyne.com
RTS Life Sciences	Automation for sample management, liquid handling, ultra-high-throughput screening and cell culture, and LIMS	Manchester, UK	www.rts-group.com
Seyonic	Nanolitre sensor-controlled liquid-handling systems; microarray spotting systems	Neuchâtel, Switzerland	www.seyonic.com
Symbiosis	Microbiology equipment to automate microbial colony counting and zone sizing	Cambridge, UK	www.symbiosis.com
Syncroscopy	Digital imaging systems for microscopy	Cambridge, UK	www.syncroscopy.com
Syngene	Systems for documenting and analysing gels	Cambridge, UK	www.syngene.com
The Automation Partnership	Automation for liquid handling, compound storage and retrieval, high-throughput screening, genomics and robotic cell culture	Royston, UK	www.automationpartnership.com
Thermo CRS	Integrated laboratory automation, magnetic bead-based separation technology	Burlington, Ontario	www.crsrobotics.com

Laboratory information management systems (LIMS)

Amersham Biosciences	Robotic workstations and lab workflow system for microarray and sequencing data	Piscataway, New Jersey	www.amershambiosciences.com
Avantium Technologies	VirtualLab integrated software, Data Analysis Pipeline software	Columbia, Maryland	www.avantium.com
CLONDIAG	PARTISAN microarray LIMS	Jena, Germany	www.clondia.com
GeneticXchange	Data access and integration software for the life sciences industry	Menlo Park, California	www.geneticxchange.com
Labtronics	Laboratory software and control systems, LIMS interfacing	Guelph, Ontario, Canada	www.labtronics.com
Labvantage	Customized LIMS	High Wycombe, UK	www.labvantage.com
L.I.M.S.	StarLIMS laboratory information management system	South Hollywood, Florida	www.starlims.com
Mitsui Knowledge Industry	LIMS with barcode	Tokyo, Japan	bio.mki.co.jp
Scinomix	Robotics and laboratory LIMS software for biotechnology	Earth City, Missouri	www.scinomix.com
Sunquest HLA	Windows-based LIMS for HLA typing	Tucson, Arizona	www.sunquest.com
Thermo Labsystems	LIMS, chromatography data-archiving and spectroscopy software	Manchester, UK	www.thermolabsystems.com

Sample storage and retrieval systems

Biophile	Ultra-low-temperature sample storage and retrieval system	Charlottesville, Virginia	www.biophileinc.com
TekCel	Automated tube and microplate management, low-temperature storage/retrieval	Hopkinton, Massachusetts	www.tekcel.com
TTP LabTech	Automated storage/retrieval, nanolitre liquid-handling, fluorescence detection	Royston, UK	www.ttplabtech.com
Universal Technology	Automated storage and retrieval systems for the pharmaceutical industry	Pittsburgh, Pennsylvania	www.univtech.com

General

ACME-Automation	Services and products for laboratory automation	Spring City, Tennessee	www.acme-automation.com
Bio-Automation	Bio-Bot linear-track robotic arm	Cheshire, Connecticut	www.bio-automation.com
Brinkmann	Laboratory instrument suppliers	Westbury, New York	www.brinkmann.com
BSI Proteomics	Proprietary system for automated protein crystallization for crystallography	Gaithersburg, Maryland	www.bsiproteomics.com
Compugen	Z4000 analysis system for complex proteomics	Tel Aviv, Israel	www.cgen.com
Hudson Control Group	Modular instrumentation for use in automated systems	Springfield, New Jersey	www.hudsoncontrol.com
Microbiology International	Automated equipment for microbiology	Frederick, Maryland	www.microbiology-intl.com
Millipore	Automated equipment for molecular biology, biochemistry, genomics and proteomics	Bedford, Massachusetts	www.millipore.com
Nunc	Microtitre plates	Roskilde, Denmark	www.nuncbrand.com
PerkinElmer Life Sciences	Automated systems for liquid handling and sample preparation	Boston, Massachusetts	lifesciences.perkinelmer.com
Phenix1	Suppliers of automated equipment and consumables for molecular biology	Hayward, California	www.phenix1.com
Pierce Chemical	Components and consumables for automated systems in protein chemistry, sample handling, immunology and chromatography	Rockford, Illinois	www.piercenet.com
RoboDesign	Automated equipment for biotech, protein structure, genomics and proteomics	Carlsbad, California	www.robodesign.com
Rohasys	Automated equipment for pretreatment, weighing, pH, conductivity, turbidity	Rijen, the Netherlands	www.rohasys.com
Sankyo Robotics	Multi-purpose robots	Boca Raton, Florida	www.sankyo.com/robotics
Structural GenomiX	Automated protein crystallography for drug discovery	San Diego, California	www.stromix.com
Syrrx	Automated protein crystallography for drug discovery	San Diego, California	www.syrrx.com
Treff	Customized plastic consumables including deep-well and PCR plates	Degersheim, Switzerland	www.treff-ag.ch
Whatman	Separation technology, microplates, Biometra instruments, and components	Maidstone, UK	www.whatman.com

● See advertisement

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager:

 Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@nature.com

US Sales Manager:

 Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe

e-mail: h.watanabe@nature.com

naturejobs

A cold shoulder for stocks

In the past, the promise of stock options was the perfect bait to lure top researchers in biotechnology out of their labs and into the corporate world as chief executives. But a recent survey of biotech chiefs in Europe suggests that this is no longer the case. With markets souring rather than soaring, cash has become a larger component in the efforts to recruit and retain biotech leadership.

The survey, by international recruitment firm Egon Zehnder in London, was aimed at finding benchmarks for the compensation offered to European chief executives. It interviewed the chiefs of 29 biotech companies across 9 countries. The average firm had 88 employees and a market capitalization of E75 million (US\$80.8 million).

The results showed a marked variation in cash salaries, before bonus, being paid — from E30,000 to E610,000, with an average of E225,000. Bonuses also varied quite widely. The average was 27% of salary, but a third of those questioned received no bonus and two received 100%. There was also a split over transport: two-thirds of the chief executives had a company car.

There was more agreement about stock packages. The average equity was 3.5% of the company's value, or about E1.52 million. When asked about their satisfaction with their packages, only 22% said that they were "just happy" with their compensation, whereas 60% said that they would like to see a higher base salary. Perhaps reflecting the recent turbulence of the stock market, only 18% wanted to receive more stock options or equity in their firm.

The upshot of this survey may be simply that the industry could use better benchmarks — and that biotech chiefs will prefer cash bonuses to stock options as incentives, until the market turns around.

Paul Smaglik
Naturejobs editor



Contents

REGIONS

Taiwan — land of career
opportunity

p672

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



Biotech vision Taiwan

A new economy with a strong biotechnology component is the Taiwanese government's ambitious vision for the future — one which it is backing with serious investment. But although it has managed to entice several big names to the island republic, there are still some interesting vacancies.

Yuan-Tseh Lee, the 1986 Nobel chemistry laureate, took over nine years ago as president of the dominant research organization, Academia Sinica. Lee was professor of chemistry at the University of California, Berkeley, and principal investigator at the Lawrence Berkeley National Laboratory before returning to his native Taiwan. Since then he has gathered veterans in a

wide range of disciplines from top Western laboratories, mainly in the United States. Yuan-Tsong Chen, for example, was professor of paediatrics and genetics at Duke University Medical Center in Durham, North Carolina, until he returned to Taiwan about a year ago to take up the directorship of Academia Sinica's Institute of

Biomedical Sciences.

Many
of the

big names drawn to Academia Sinica — based in the capital, Taipei — are Taiwanese-born scientists who wanted to return home and do some good for their native land. "They need my expertise here," says John Yu, a stem-cell researcher who left the Scripps Research Institute in La Jolla, California, in August 2002 and now heads the zoology institute.

Universities, too, have begun to tap overseas resources. Last year the National Tsing Hua University in Hsinchu, just south of Taipei, brought in astronomer Frank Hsia-San Shu from the University of California, Berkeley, as president. He hopes that his long years of experience in the California university system should help him carry out an administrative shake-up.

The government's investment in science has enhanced Taiwan's attractions as a research destination over the past few years. An extra 10% on the science budget — despite overall spending cuts — has paid for huge national projects, in pharmaceuticals and biotechnology in 2000 and genomic medicine in 2002. It has also provided new facilities, equipment and support that Chen, Yu and many of their colleagues say rival the best available anywhere.

PERSONNEL SHORTAGE

Despite the infrastructure investment, Taiwan is still desperately short of personnel. Recruiting high-level researchers is often a question of finding overseas Taiwanese who want to help their homeland. But the huge number of Taiwanese students who used to go abroad for education has shrunk, leaving a smaller reservoir of expatriates to draw on. Many of the best who stay end up forsaking research for a job in one of the industrial science parks.

The flow of postdoctoral students from mainland China has also been stemmed, as opportunities in

Nobel laureate Yuan-Tseh Lee has helped to galvanize research efforts at Taiwan's Academia Sinica.



Beijing and Shanghai grow. "Many people have a grant that includes a postdoc, but can't find one," says James Shen, director of Academia Sinica's Institute of Molecular Biology.

Efforts are being made to improve graduate education at seven of Taiwan's leading universities, but this will be slow in producing the necessary researchers. Academia Sinica is pushing ahead with an international graduate school to bring in talent from abroad. But although it provides great opportunities for aspiring researchers, it is still small (see *Nature* **421**, 464; 2003).

This leaves numerous opportunities for researchers at all levels. Taiwan's scientific leaders pin their hopes on the investment that has been made in infrastructure, such as adding a set of nuclear magnetic resonance machines and a new mouse facility at Academia Sinica. "Once these projects become visible it will help bring people here," says Shen.

OUTSIDER'S PERSPECTIVE

Taiwanese-born returnees are not the only researchers who are now enjoying outstanding facilities, especially ample computer-processing power — a benefit of being on an island that made its fortune in microelectronics. "It would take me a year to do what I can do in a day here," says astrophysicist Alexandr Serebryanskiy, a postdoctoral student at National Tsing Hua University who came from Uzbekistan two years ago.

"I wouldn't know where to look for these kinds of facilities at home," agrees Jon Wright, a British postdoctoral student who moved to Academia Sinica in 1996. "And if I found them, I would probably have to share with many others."

Western researchers generally find that Taiwan's research environment fosters a fruitful, open exchange of ideas in which language is not a problem. In everyday life, too, English works well enough at supermarkets and hospitals for researchers and their families to feel comfortable without having to learn Chinese.

But an inability to speak the language does have its pitfalls. Conference announcements, travel-grant materials and other notices are often available only in Chinese, which can leave foreign researchers feeling isolated. Those planning to bring families may balk at the exorbitant price of Taipei's main international schools, although some have found affordable international or bilingual schools in Hsinchu and the outer suburbs of Taipei.

If Taiwan wants to compete for international talent, it may need to raise salaries, which have been held down by what one Academia Sinica researcher calls an

this salary sacrifice. A new programme to be started by the National Science Council last August will offer foreign recruits named as "natural-science research professors" salaries in line with those of their home institutions.

Although a pay rise would be useful, many foreign researchers are more concerned about job security: they say that difficulty in getting a full-time research position makes for an unstable life. Many are left in one-year postdoctoral or short-term visiting scientist positions. This sends a mixed message, says one Academia Sinica postdoctoral student. Although the infrastructure and grant support are attractive, the lack of long-term opportunities gives pause. "It is difficult to think of making a career here," the student says.

There are signs of change. Frank Shu is keen to make the National Tsing Hua University truly international, with some 40 staff teaching in English. Eventually, he hopes to have all graduate education in English. "We need more diversity in Asia," he says.

He also hopes to restructure the "too restricted" salary levels, to open up greater funds for younger researchers. "You can't let promising young researchers waste their first five years begging for grants," Shu says.

And the investment continues. A biomedical park, going up in the outer suburbs of Taipei near Hsinchu, will have a medical centre affiliated with Taipei's National Taiwan University and research centres planned with Academia Sinica's help. When the medical centre is opened in 2004, it will only be a 20-minute ride on a new bullet train from Taipei. And inroads into biotechnology are also being established (see 'Foundations for biotech', below).

The pieces are in place to make Taiwan a tempting research destination for both natives and foreigners, says biochemist and Academia Sinica vice-president Sunney Chan. When Chan returns to the California Institute of Technology this August, he will leave Taiwan with several initiatives, including a better peer-review system and a national project in genomics. "The system and facilities are there," Chan says. "Now we just have to ask the right scientific questions."

David Cyranoski is *Nature's* Asia-Pacific correspondent.



Back home: Yuan-Tsong Chen recently made the switch from the United States to Academia Sinica.

James Shen hopes that Taiwan's investment will lure many researchers to the country.



"overly democratic" impulse to give all researchers the same wage. Yuan-Tseh Lee himself set up a Foundation for the Advancement of Outstanding Scholarship, to ease the problem by providing stipends for about 100 researchers. But many people still have to accept a much lower salary than they have been used to.

University and government leaders are taking measures to correct

Foundations for biotech

One of the most promising starting points for growth in biotechnology in Taiwan is likely to be the Biomedical Engineering Center (BMEC) in the Hsinchu Science-based Industrial Park. Founded three years ago as part of the Industrial Technology Research Institute, the organization behind Taiwan's rise in microelectronics, the BMEC has about 400 researchers, including 50 postdoctoral students.

Following the lead of the BMEC's parent organization, successful technologies, and the teams that develop them, will be spun off to create companies. Apart from more conventional projects in tissue engineering and bioinformatics, the BMEC is developing an artificial 'nose' — a chip with peptides that simulate the protein structure of olfactory molecules and can detect liver cirrhosis and other diseases. D.C.